Paralogy in Cladistic Biogeography and Analysis of Paralogy-Free Subtrees

G. NELSON[1] AND P. Y. LADIGES[2]

CONTENTS

[1] Curator, Department of Herpetology and Ichthyology, American Museum of Natural History.
[2] Professor and Head of School, School of Botany, The University of Melbourne, Parkville, Victoria, Australia.

ABSTRACT

Geographic paralogy—analogous to the molecular phenomenon—has gone unrecognized in cladistic biogeography. It is evidenced by duplication or overlap in geographic distribution of taxa related by a particular node of a cladogram of organisms. Geographically paralogous nodes increase basally, therefore nonrandomly, in cladograms generally, such that most nodes of complex cladograms of organisms are geographically paralogous. A novel algorithm, implemented in a preliminary MS-DOS program, reduces a more or less complex cladogram of organisms to one or more subtree (area cladogram) that is paralogy free. Subtree analysis of a number of published studies indicates that geographic data associable with informative nodes of such subtrees appear to be the only data relevant to cladistic biogeography; such data, represented as either components or three items in a matrix for parsimony analysis, are found to be remarkably consistent; most geographic inconsistency previously noted in cladistic biogeography, through parsimony analysis of matrices of geographic data, is merely the effect of paralogy.

INTRODUCTION

In cladistic biogeography, nodes of a cladogram of organisms are potentially informative about relationships among geographic areas occupied by the organisms. Cladistic biogeographers associate geographic data with each node, and combine and interpret the data of all nodes of one or more cladogram of organisms. Availability of parsimony programs encouraged their use for such purposes. The programs, designed to find one or more tree that best fits a particular sample of data, require that data be organized in binary form (zeros and ones), arranged in a matrix of rows corresponding to areas, and columns (characters) corresponding to nodes. This approach developed various ways to associate nodes and geographic data, and various ways to represent data in a matrix (summaries in Humphries and Parenti, 1986; Humphries et al., 1988; Humphries, 1992; Legendre, 1990). We consider two general notions of data for cladistic biogeography, apply them to two benchmark studies, those of Brundin (1966) and Mayden (1988), and suggest a complementary notion (subtree algorithm) that offers hope of better results. We illustrate this possibility through analysis of these two studies and also of some of those reviewed by Craw (1989), Page and Lydeard (1994), and Morrone and Carpenter (1994).

METHODS

COMPUTER PROGRAMS

Programs used in analyses include Hennig86 (Farris, 1988), PAUP (Swofford, 1993), and those in the current TAX package (Nelson and Ladiges, 1995). Hennig86 and PAUP are programs used for parsimony analysis of matrices. Hennig86 was used for matrices with fewer than 1000 characters and for search for minimal trees (see below); and PAUP, for matrices with more than 1000 characters. The TAX package includes programs (TAX, TAS) used for preparation of three-item matrices for analysis by Hennig86 and PAUP. The package includes also a preliminary program (TASS) for subtree analysis (used to enumerate subtrees and prepare both component and three-item matrices for subtrees); and a utility program (TAXUTIL) for analysis of tree files (used to enumerate nodes per tree in a file and three-item statements per tree, and to select trees from a file).

SUBTREE ALGORITHM

As described and exemplified below, and implemented in the program TASS, the subtree algorithm builds subtrees starting at each terminal node and progressing to the base of a cladogram of organisms. A node (taxon) that relates organisms that, as different taxa, do not overlap in geographic distribution is associated with the nonoverlapping geographic data. A node (taxon) that relates organisms that, as different taxa, overlap is deemed paralogous and is not generally associated with geographic data, except in the following case: if a node leads directly to one or more terminal taxon that is geographically

Fig. 1. Columns: Cladograms of organisms — Component matrices — Three-item matrices

Example 1

Cladogram: ⊢0 ⌐1 — A / B / C

Component matrices	nodes: 0	1	Three-item matrices	1
A	1	0	A	0
B	1	1	B	1
C	1	1	C	1

Example 2

Cladogram: ⊢0 ⌐1 — A / B ⌐2 C1 / C2

Component matrices	nodes: 0	1	2	Three-item matrices	1	1
A	1	0	0	A	0	0
B	1	1	0	B	1	1
C	1	1	1	C	1	1

Example 3

Cladogram: ⊢0 ⌐1 — A1 / A2 ⌐2 B / C

Component matrices	nodes: 0	1	2	Three-item matrices	2	2
A	1	1	0	A	0	0
B	1	1	1	B	1	1
C	1	1	1	C	1	1

Example 4

Cladogram: ⊢0 ⌐1 — C2 / A ⌐2 B / C1

Component matrices	nodes: 0	1	2	Three-item matrices	1	2
A	1	1	0	A	1	0
B	1	1	1	B	1	1
C	1	1	1	C	0	1

Example 5

Cladogram: ⊢0 ⌐ C2 / C3 ⌐1 A ⌐2 B / C1

Component matrices	nodes: 0	1	2	Three-item matrices	1	1	2
A	1	1	0	A	1	1	0
B	1	1	1	B	1	1	1
C	1	1	1	C	0	0	1

Example 6

Cladogram: ⊢0 ⌐1 ⌐4 A1 / B / C1 / ⌐2 A2 / C2 / ⌐3 A3 / C3

Component matrices	nodes: 0	1	2	3	4	Three-item matrices	1	1	2	3	4	4	4
A	1	1	1	1	0	A	1	1	1	1	0	0	0
B	1	1	0	0	1	B	1	1	0	0	1	1	1
C	1	1	1	1	1	C	0	0	1	1	1	1	1

Fig. 1. Six cladograms (Examples 1–6) of organisms in areas A, B, and C, with data for nodes represented in component and three-item matrices (all-zero outgroup assumed).

widespread, and part of that distribution overlaps with that of another taxon, or taxa, then the widespread distribution is reduced to the nonoverlapping geographic element. An example is that of five taxa with distributions in areas A, AB, B, BC, D, and related by the cladogram ((A AB)(B BC))D. All three informative nodes of the cladogram are paralogous in the strict sense, yet one node evidently relates areas A and B more closely than to D, and another node evidently relates areas B and C more closely than to D. The algorithm reduces the widespread distributions AB and BC to B and C, respectively, in accordance with assumption 2 (Nelson and Ladiges, 1991b). The cladogram with reduced distributions yields two subtrees, (AB)D and (BC)D, which combine as (ABC)D. The same result is obtained from the cladogram ((A AB)(B BC))BD.

MINIMAL TREES

A matrix with missing data invites existing parsimony programs to find trees that have shortest length but are overresolved for the data of the matrix (Platnick et al., 1991; Nelson and Ladiges, 1993). Current parsimony programs save such trees in computer memory if the trees do not conflict with data, even though most of their nodes might be unsupported by data. Saving overresolved trees is a defect of current programs, but the defect can be overcome through search for a least-resolved (minimal) tree. Search for minimal trees was conducted with the xx function of Hennig86. After entering the appropriate matrix and the appropriate tree, nodes were collapsed, and areas moved to a more basal position in the tree, one at a time and in combination on a trial-and-error basis, in a search for the least resolved tree with shortest length. A tree was judged minimal when further collapse increased tree length.

GEOGRAPHIC DATA (FIG. 1)

There are two basic notions about geographic data associable with nodes of a cladogram relating organisms. Both notions concern data organized as 0-entries and 1-entries in a matrix. We refer to the data of these notions as "components" (Nelson and Platnick, 1981: 169) and "three items" (Nelson, 1992; Nelson and Ladiges, 1993) and illustrate some differences between them in six examples (fig. 1).

EXAMPLE 1

Consider an example of organisms that live in geographic areas A, B, and C. Organisms of areas B and C (node 1) are related more closely than to organisms of area A. Both nodes (0 and 1) have components associable with them, whereas only one node (1) has three items.

The rationale for components is to follow each node to all branch tips to which it leads and to sum the data at the tips. Node 0 leads to three tips with data A + B + C; hence for node 0 of the matrix, each area receives the value 1. Node 1 leads to two tips with data B + C; hence for node 1 of the matrix, area A receives the value 0, and areas B and C receive the value 1.

The rationale for three items is to see each

node as a relation (connection) between the branches (and their tips) to which it leads—relating these branches more closely than to other branches (and tips) of the tree. If there were another area, e.g., area D, at the tip of a more basal branch, then node 0 would relate area A and areas B and C more closely than to area D. Because no such area D is included in the tree, node 0 does not function as a relation (connection) with that significance, and node 0 is without data. In contrast, node 1 relates areas B and C more closely than to areas at the tips of other branches. Here the only such area is A. Hence for node 1 of the matrix, area A receives the value 0, and areas B and C receive the value 1.

With an all-zero outgroup, parsimony analysis of each matrix yields one tree, A(BC) with consistency index (ci) 100.

EXAMPLE 2

Organisms occur in areas A, B, and C, with two sorts of organisms (C1 and C2) in area C. Two sorts of organisms living in area C are related (node 2) more closely than to organisms of areas A and B, and are related (node 1) more closely to organisms of area B than to organisms of area A. All three nodes have component data. Only node 1 has three-item data, but it has twice as much data as node 1 of Example 1 (see above). Parsimony analysis of each matrix yields one tree, A(BC), with ci 100.

The rationale for components is the same as that of Example 1, with the added factor of multiple occurrence, or redundancy, in area C. Node 0 leads to four branch tips with data A + B + C + C. Redundancy in the data is simply eliminated, leaving the data for node 0 as ABC. Similarly, node 1 leads to three branch tips, with data B + C + C, reducing to BC. Node 2 leads to two branch tips with data C + C, reducing to C.

The rationale for three items sees geographic data in a different way. Node 0 is without data (as in Example 1). Node 1 relates area B and area C (C1) more closely than to area A, and again relates area B and area C (C2) more closely than to area A; hence in the matrix, two columns (characters) are required for node 1. Node 2 cannot logically relate area C (C1) and area C (C2) more closely than to areas A and B; hence, node 2 is seen as without data. Relationship is a con-

nection between different areas; to claim that an area is related, or connected, to itself more closely than to another area is logically absurd.

EXAMPLE 3

Organisms occur in areas A, B, and C, with two sorts of organisms (A1 and A2) in area A. Organisms of areas B and C are related (node 2) more closely than to organisms of area A, but are related (node 1) to one sort of organisms (A2) of area A more closely than to the other sort (A1). All three nodes have component data. Only node 2 has three-item data, but it has twice as much data as node 1 of Example 1 (see above). Parsimony analysis of each matrix yields one tree, A(BC), with ci 100.

The rationale for components is similar to that of Example 2. Node 0 leads to A + A + B + C, reducing to ABC, and so on. The rationale for three items sees node 1 as without data: node 1 cannot logically relate area A (A2) and areas B and C more closely than to area A (A1). Node 2 relates areas B and C more closely than to area A (A2), and again relates areas B and C more closely than to area A (A1); hence in the matrix, two columns (characters) are required for node 2.

EXAMPLE 4

Organisms occur in areas A, B, and C, with two sorts of organisms (C1 and C2) in area C. Organisms of one sort (C1) of area C and those of area B are related (node 2) more closely than to those of area A and to those of the other sort (C2) of area C, and are related (node 1) more closely to those of area A than to the other sort (C2) of area C. All three nodes have component data (with exactly the same characters as Example 3). Only nodes 1 and 2 have three-item data. Parsimony analysis of the component matrix yields one tree, A(BC), with ci 100. Parsimony analysis of the three-item matrix yields two trees, (AB)C and A(BC), with ci 66.

The rationale for components is similar to that of preceding examples. In the rationale for three items, node 1 relates areas A and B more closely than to area C (C2), but cannot logically relate areas A and C (C1) more closely than to area C (C2). Node 2 relates areas B and C (C1) more closely than to area A,

but cannot logically relate areas B and C (C1) more closely than to area C (C2). Three-item data conflict, (AB)C versus A(BC), and component data do not. Parsimony analysis of the three-item matrix does not resolve the conflict because there is no preponderance of one pattern. Both trees, (AB)C and A(BC), have length 3 for the three-item matrix.

EXAMPLE 5

Organisms occur in areas A, B, and C, with three sorts of organisms (C1, C2, C3) in area C. Organisms of area B and one sort (C1) of area C are related (node 2) more closely than to those of area A and to the other sorts (C2 and C3) of area C, and are related (node 1) more closely to those of area A than to the other sorts (C2 and C3) of area C. All three nodes have component data (with exactly the same characters as Examples 3 and 4). Only nodes 1 and 2 have three-item data. Parsimony analysis of the component matrix yields one tree, A(BC), with ci 100. Parsimony analysis of the three-item matrix yields one tree, (AB)C, with ci 75.

The rationale for components and the rationale for three items are similar to those of the preceding examples. In this example, three-item data conflict, (AB)C versus A(BC), and component data do not. Parsimony analysis of the three-item matrix resolves the conflict because there is a preponderance of one pattern, (AB)C. Tree (AB)C has length 4, and tree A(BC) has length 5 for the three-item matrix.

EXAMPLE 6

Organisms occur in three areas, with three sorts of organisms (A1, A2, A3) in area A, and three sorts (C1, C2, C3) in area C. All five nodes have component data. Only nodes 1–4 have three-item data. In the three-item matrix, there are two characters for node 1, (A1B)C2 and (A1B)C3; one character each for nodes 2 and 3, (A2C2)B and (A3C3)B, respectively; and three characters for node 4, A1(BC1), A2(BC1), and A3(BC1).

Parsimony analysis of the component matrix yields one tree, (AC)B length 6, ci 83, retention index (ri) 66; with uninformative characters (0, 1) inactive, length 4, ci 75, ri 66. Parsimony analysis of the three-item ma-

trix (uniformly or fractionally weighted) yields one tree, A(BC), length 11, ci 63, ri 47.

SIGNIFICANCE OF EXAMPLES 1–6

These six examples demonstrate differences between component and three-item data. Noteworthy is the relative insensitivity of component data to differences among cladograms of organisms. In Examples 1–5, the single informative character of each component matrix is exactly the same, whereas the informative character(s) of the three-item matrix are generally different. In Examples 3–5, the entire component matrix is exactly the same, whereas the three-item matrix is different. In Examples 4 and 5, the component matrix contains no conflict, whereas the three-item matrix does, and parsimony analysis of the matrices yields different trees. In Example 6, both types of matrices contain conflict, but the conflict is differently represented in the matrices, and parsimony analysis of the matrices yields different trees.

It is pointless to argue that different cladograms (Examples 3–5) are exactly represented by the same matrix. Rather, different cladograms imply different matrices for their exact representation. From these examples it seems that three-item data are more exact than component data. It does not follow, however, that geographic results of parsimony analysis of the three-item matrices are, therefore, generally better than results of parsimony analysis of the component matrices. It is enough to note that three-item data include variability that component data sometimes exclude. In what follows, the cause and significance of this variability are considered in detail.

AUSTRAL MIDGES (Brundin, 1966)

A publication significant in the history of cladistic systematics and biogeography is that of Brundin (1966). His monograph of species of certain groups of midges of the Southern Hemisphere contains a general discussion of cladistic principles, their application to midges, and conclusions about the history in time and space of these organisms. His monograph began the cladistic approach to biogeography. More than any previous publication, it influenced subsequent develop-
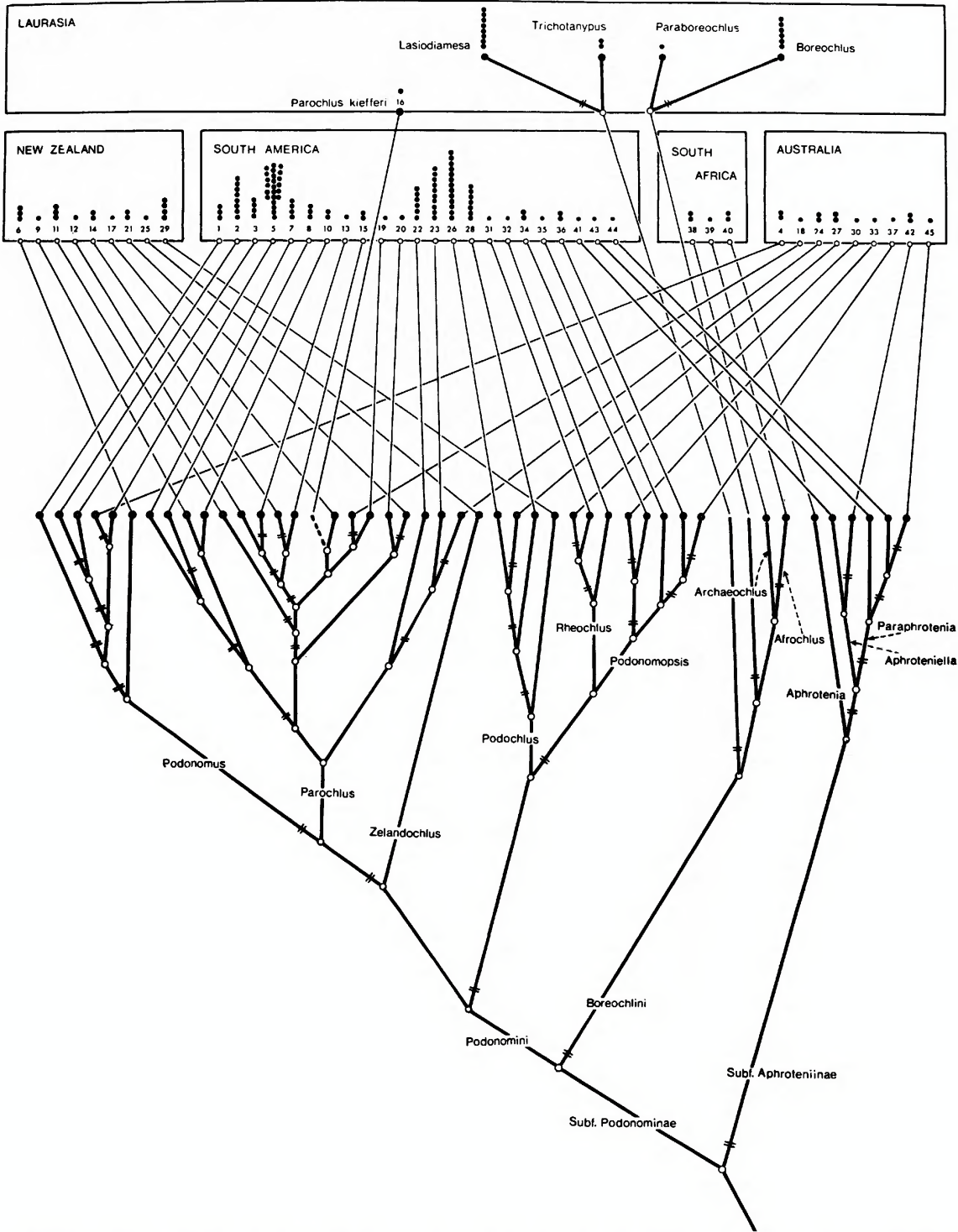
Fig. 2.   Relationships of midges of subfamilies Podonominae and Aphroteniinae (after Brundin, 1966: fig. 634, modified from 1965: fig. 2; 1967: fig. 2).
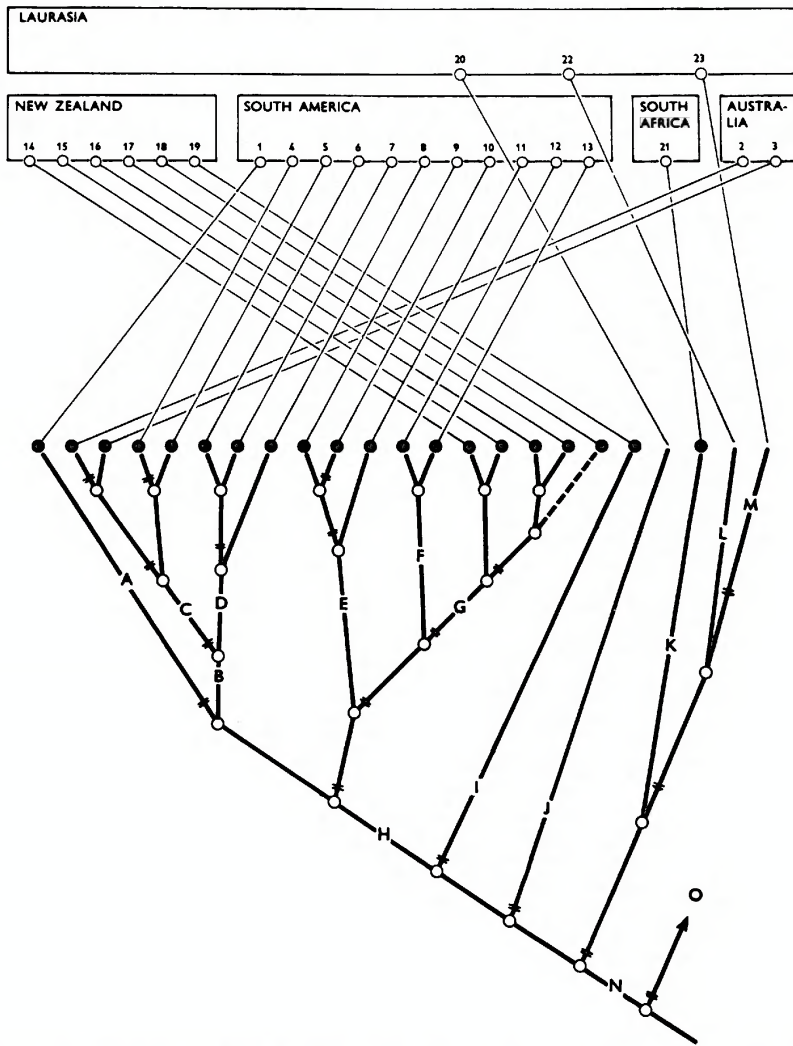
Fig. 3. Relationships of midges of subfamily Diamesinae (after Brundin, 1966: fig. 635).

ment of cladistic systematics and biogeography.

In Brundin's monograph, there are cladograms of two groups of midges comprising three subfamilies and 70 terminal taxa (figs. 2, 3), each a species or species group, and a summary of their geographic relationships (fig. 4). Brundin later (1970: fig. 1; 1972: fig. 4; 1975: fig. 1; 1988: fig. 11.2) offered a second summary (fig. 5). Still later (1972: figs. 5-6; 1975: figs. 2-3) he offered a third summary (fig. 6).

In Brundin's first summary, there are two geographic patterns (fig. 7, P1 and P2). In P1, southern South America relates to New Zea-

land, and these two areas relate to southern Africa; in P2, southern South America relates to southeastern Australia, and these two areas relate to southern Africa. If P1 and P2 are combined, then three of the four areas (South America, New Zealand, Australia) relate among themselves, and as a group relate to southern Africa (fig. 7, P1 + 2; Nelson and Ladiges, 1991b: table 6, Example 2).

Brundin's second summary (fig. 7, S2) shows that Australia and South America 2 (node 4) are related more closely than to South America 1, these three areas are related (node 3) more closely than to New Zealand, these four areas are related (node 2) more closely
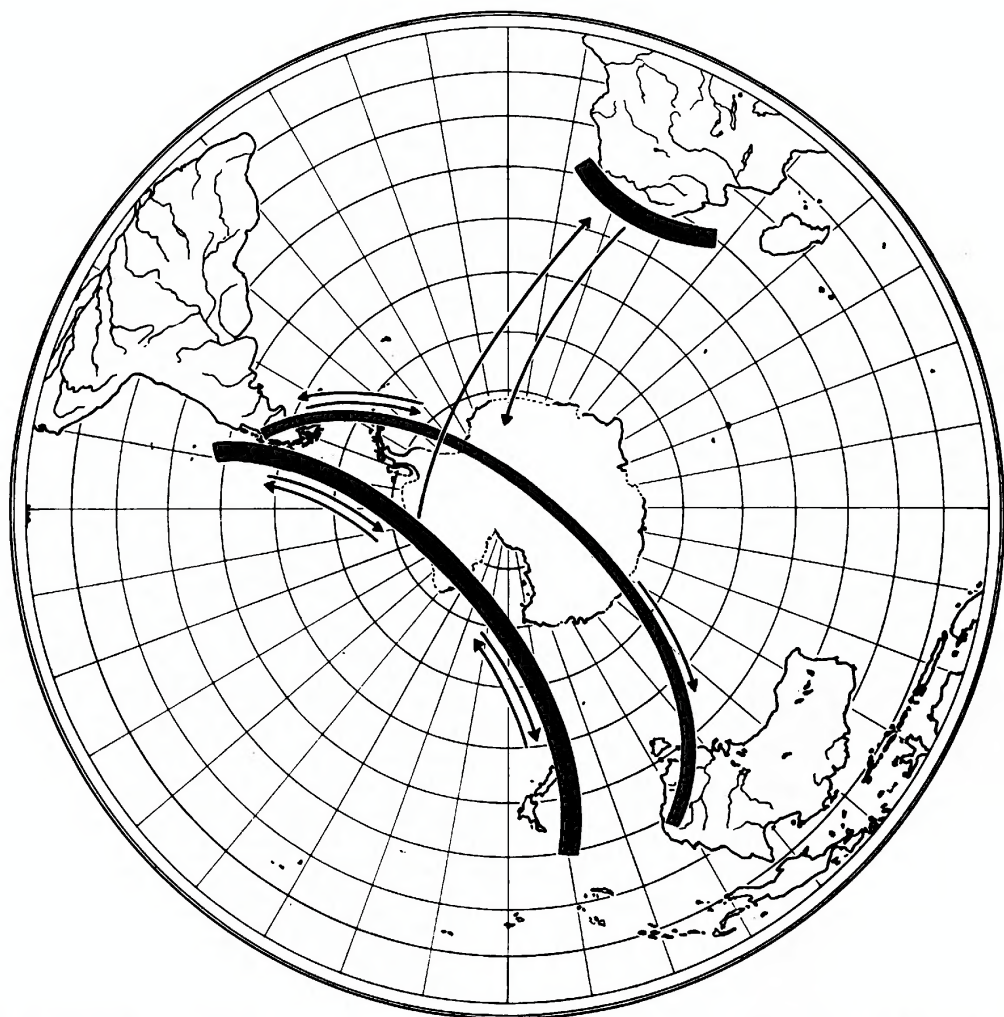
Fig. 4. "Main pattern of transantarctic relationships, as evidenced by chironomid midges. The black arcs connect the inferred Mesozoic main nodes of evolution and dispersal of the paleoaustral element. The arrows indicate directions of dispersal before the disruption of Gondwanaland" (after Brundin, 1966: fig. 636).

←
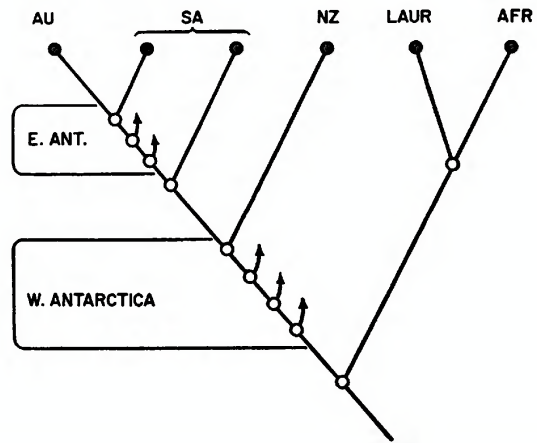
Fig. 5. "The connection between phylogenetic relationship, relative age, and geographical distribution in cold-adapted chironomid midges of austral origin. Circles with attached arrows indicate the multiple occurrence of accordant transantarctic connections within a monophyletic group. The different evolutionary and biogeographical role played by East and West Antarctica after the separation of South Africa from the other southern lands in the Upper Jurassic is also indicated" (after Brundin, 1970: fig. 1; 1972: fig. 4; 1975: fig. 1; 1988: fig. 11.2).
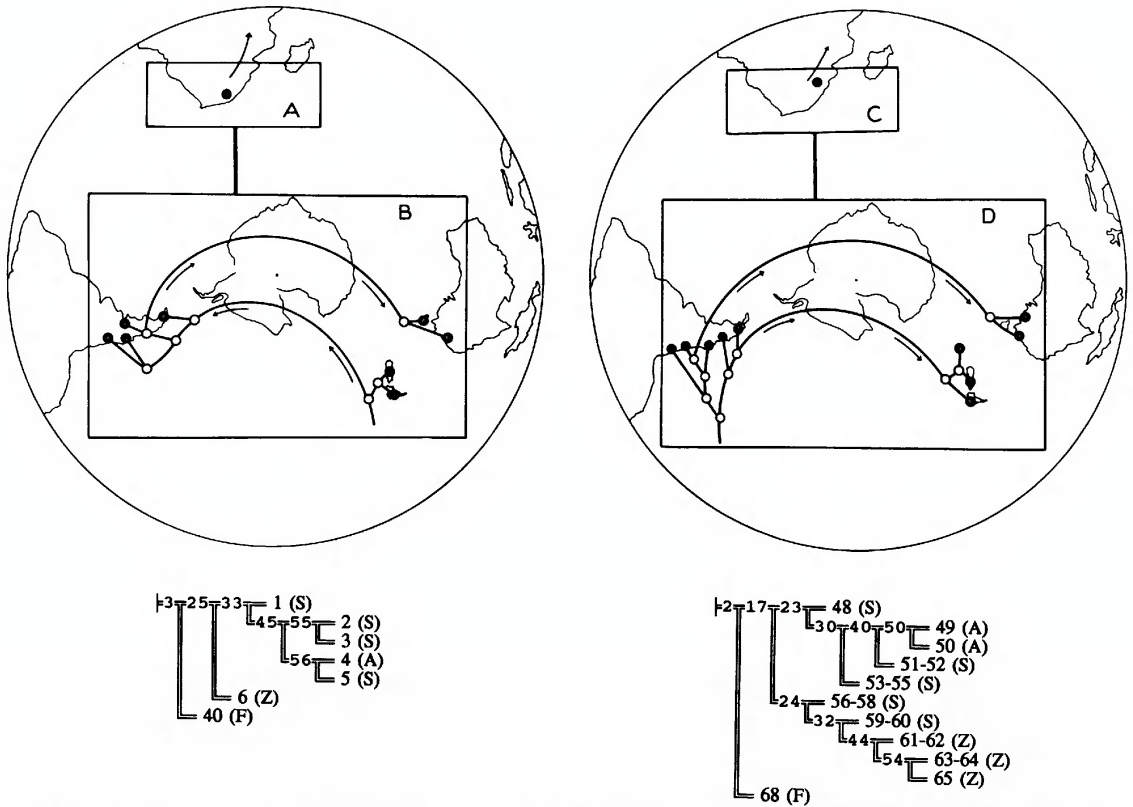
Fig. 6. "Circum-Antarctic distribution and inferred transantarctic relationships." **Above:** left (subfamily Podonominae), "A: the tribe Boreoclini; B, the tribe Podonomini. The phylogenetic diagram (within the frame of 'B') refers to the situation in the genus *Podonomus*, where the species group of New Zealand is plesiomorphic and the species group of Australia is apomorphic in relation to the corresponding sister groups in South America"; **right** (subfamily Diamesinae), "C: the Diamesae; D: the Heptagyiae. The phylogenetic diagram (within the frame of 'D') refers to the situation in the tribe Heptagyini, where the group of New Zealand (genus *Maoridiamesa*) and the group of Australia (the *tonnoiri* group of the genus *Paraheptagyia*) are both apomorphic in relation to the corresponding South American sister group" (after Brundin, 1972: figs. 5, 6; 1975: figs. 2, 3). **Below,** portions of combined cladograms (fig. 8) represented above.

than to Africa and Laurasia, which are related (node 1) among themselves. Laurasia aside, Brundin's second summary, containing two nodes (3 and 4) that relate Australia and South America more closely than to New Zealand, differs from the combination of the two patterns. Even so, the second summary might be more accurate than the two patterns.

PARSIMONY ANALYSIS: NODES OF COMBINED MIDGE CLADOGRAMS (FIG. 8)

Brundin's two cladograms (figs. 2, 3) can be combined via a basal node as one cladogram of 69 nodes and 70 terminal taxa (fig. 8), distributed in Australia(A), Africa (F), Laurasia (L), South America (S), and New Zealand (Z). Table 1 is the component matrix for all 69 nodes, most of which are variably redundant (tables 2, 3). Parsimony analysis of this matrix yields one tree (fig. 7, Comp), similar to Brundin's second summary (see above and fig. 7, S2).

Three-item analysis of nodes of the combined cladograms (fig. 8) yields 17,051 statements. Parsimony analysis of a uniformly weighted matrix yields one tree (fig. 7, UW). Parsimony analysis of a fractionally weighted matrix ($\times 10$) yields one tree (fig. 7, FW $\times 10$).

Results of parsimony analysis of the matrices for the nodes of the combined clado-
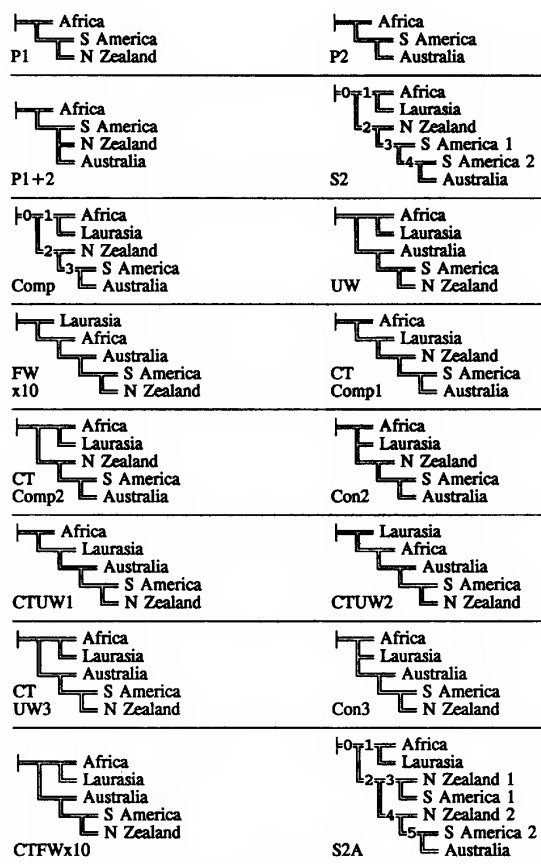
Fig. 7. Relationships of austral areas, determined by various means. P1 and P2, the two patterns of Brundin's first summary (fig. 4). P1 + 2, combination of patterns P1 and P2. S2, Brundin's second summary (fig. 5). Comp, result (one tree) of parsimony analysis of component matrix (table 1) for 69 nodes of combined cladograms (fig. 8), and also of component and three-item matrices (tables 4, 5) for 19 nodes of 16 subtrees (fig. 9); for matrix of table 1, length 79, ci 87, ri 84—with uninformative characters rendered inactive (characters 0–3, 11, 21, 31, 41–44, 50–55, 57, 61, 67), the reduced matrix yields the same tree, length 59, ci 83, ri 84; for matrices of table 4, length 19, ci 100, ri 100; for uniformly weighted matrices of table 5 (44 statements), length 44, ci 100, ri 100; for fractionally weighted matrices of table 5, length 44 × any factor, ci 100, ri 100. UW, result (one tree) of parsimony analysis of uniformly weighted three-item matrix (17,051 statements) for 52 informative nodes of combined cladograms (fig. 8), length 26441, ci 65, ri 50. FW × 10, result (one tree) of parsimony analysis of fractionally weighted (× 10) three-item matrix for 52 informative nodes of combined cladograms (fig. 8), length 260722, ci 65, ri 50. CTComp1 and CTComp2, results (two trees) of parsimony analysis of component matrix (table 6) for 23 nodes of combined subtrees (fig. 10), length 30, ci 76, ri 72. Con2, strict consensus of two trees (CTComp1 and CTComp2). CTUW1, CTUW2, CTUW3, results (three trees) of parsimony analysis of uniformly weighted three-item matrix (1682 statements) for 22 informative nodes of combined subtrees (fig. 10), length 2564, ci 65, ri 47. Con3, strict consensus of three trees (CTUW1, CTUW2, CTUW3). CTFW × 10, result (one tree) of parsimony analysis of fractionally weighted (× 10) three-item matrix for 22 informative nodes of combined subtrees (fig. 10), length 24400, ci 66, ri 48. S2A, Brundin's second summary, modified to conform exactly with relationships of austral areas as he saw them.

grams differ (fig. 7, Comp, UW, FW × 10). These different results for component and three-item data (with uniform and fractional weighting) indicate that the data of the combined cladograms are ambiguous. The source of this ambiguity is geographic paralogy.

GEOGRAPHIC PARALOGY

Paralogy is a term used by molecular biologists to refer to comparison (or the relation) between copies of the same gene within a genome. Geographic paralogy, analogous with the molecular phenomenon (Nelson and Ladiges, 1991b: 481; Page, 1993), is duplication or overlap in geographic distribution among related taxa (hereinafter the term par-

alogy refers to geographic paralogy). A cladogram node is paralogous when it relates organisms with geographic distributions that overlap to any degree, and such distributions are themselves paralogous (cf. the "redundant" nodes of Page, 1988: 269, 1994: 65). In four of the five examples above (fig. 1), there is at least one paralogous node: node 2 of Example 2, node 0 of Example 3, node 0 of Example 4, node 0 of Example 5. In Brundin's combined cladograms (fig. 8), there are many (47) paralogous nodes (table 2).

There are several possible causes of paralogy in the sense of overlapping geographic distributions of taxa: tectonics, dispersal, sympatric speciation, mistaken relationships among organisms, imprecise characteriza-
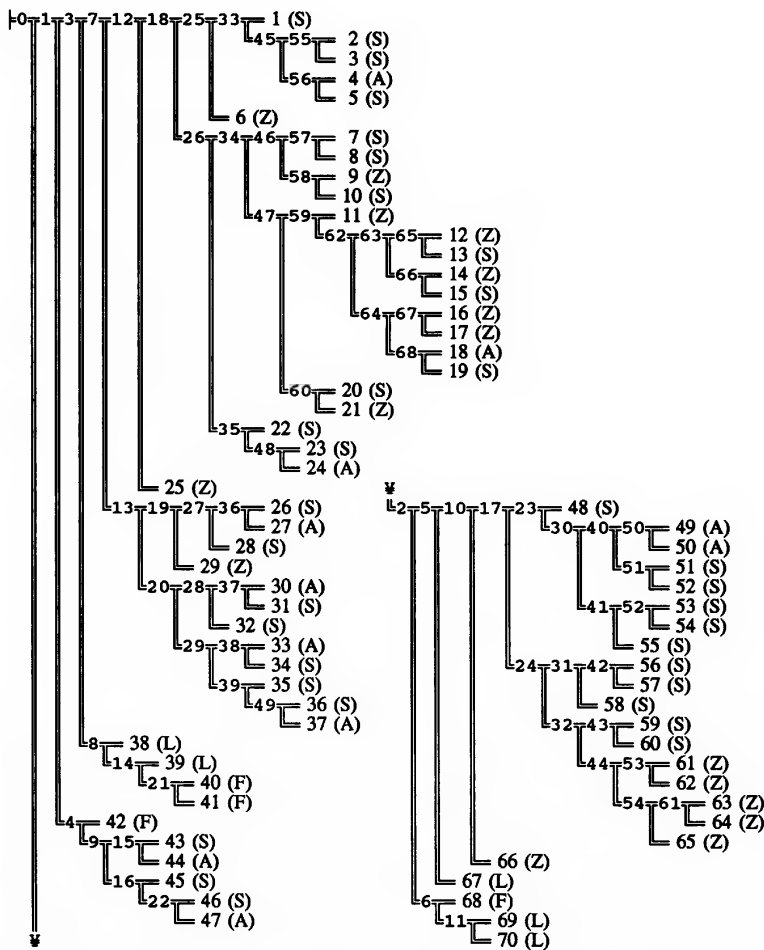
Fig. 8. Combination of Brundin's two cladograms (figs. 2, 3). *Paroclus kiefferi* (terminal taxon 16 with a Laurasian distribution), which Brundin arbitrarily placed as the sister-species of *P. maorii* of New Zealand (cf. Brundin, 1963: fig. 4), is listed as a New Zealand taxon; elsewhere (Brundin, 1965: fig. 2; 1967: fig. 2), *P. kiefferi* is shown as sister to a group including terminal taxa 17–19, and still elsewhere Brundin (1976: 148) states that *P. kiefferi*, once seen with "a clearly secondary occurrence in the Holarctic," is now part of a "changed . . . picture" (cf. Brundin, 1981: 117–118). Terminal taxa 1–47 are as in figure 2 (40–47 here are 38–45 in fig. 2). Terminal taxa 48–70 are as in figure 3 (1–23 in fig. 3).

tion of geographic areas, and so on. It is evident, however, that in cladograms of many taxa paralogy generally increases toward the base of the cladogram, such that, beyond a certain point, all more basal nodes are likely to be paralogous. Such is evident in Brundin's combined cladograms (fig. 8; see below).

SUBTREE ALGORITHM AND COMBINED MIDGE CLADOGRAMS (FIG. 9)

If within cladograms generally, paralogous nodes are nonrandomly distributed, then a

data matrix might be improved if paralogous nodes were identified so that no geographic data were associated with them. We suggest an algorithm for that purpose. The objective of the algorithm is to reduce a cladogram, such as that of Brundin's combined cladograms, to one or more subtree that is paralogy free. The algorithm begins with a list of terminal nodes that are nonparalogous and geographically informative. The algorithm builds subtrees, starting with each terminal node, progressing basally node by node, and incorporating nonparalogous geographic data for

each node while ignoring any paralogous data (as indicated by overlap of geographic areas). Of the six examples above (fig. 1, Examples 1–6), each reduces to one and the same subtree, A(BC).

Brundin's combined cladograms (fig. 8) reduce to 16 subtrees (fig. 9), which together include 19 informative nodes. Of the 69 nodes of the combined cladograms (fig. 8), 47 are paralogous. Three nodes (3–5) are in themselves uninformative, but each of the three functions as a basal node of more than one subtree (table 2).

PARSIMONY ANALYSIS OF NODES

INDIVIDUAL MIDGE SUBTREES (FIG. 9)

Nodes of individual subtrees may be represented in either a component or a three-item matrix. For the 19 nodes of the subtrees, a component matrix (table 4, left) has 19 characters. Parsimony analysis of this matrix yields one tree, length 19, ci 100, ri 100. The same result is obtained for a matrix of 16 characters (including three multistate characters), each representing one subtree (table 4, right). Three-item analysis of the 19 nodes yields 44 statements (table 5). Parsimony analysis of a uniformly weighted matrix yields one tree, length 44, ci 100, ri 100. Parsimony analysis of a fractionally weighted matrix (any factor) yields one tree, length 44 × any factor, ci 100, ri 100. However represented, the data yield one and the same tree (fig. 7, Comp)— similar to Brundin's second summary (fig. 7, S2).

For nodes of subtrees, parsimony analysis of both types of matrices (component and three-item) yields the same tree with 100% consistency, suggesting that geographic paralogy is the sole cause of the variable results of parsimony analysis of matrices (component and three item) of nodes of combined cladograms (see above).

Parsimony analysis of matrices for the nodes of the individual subtrees yields a tree (ci 100) the same as those (ci 83–87) yielded by parsimony analysis of the component matrix for nodes of combined cladograms (see above), suggesting that in this case component data—in contrast to three-item data—reduce the effects of paralogy within combined cladograms. Such is not a general prop-

TABLE 1
Component Matrix for 69 Nodes of Brundin's Combined Cladograms (fig. 8)[a]

	Nodes of Combined Cladograms
	0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2 3 3 3 3 3
	0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4
OG	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
A	1 1 1 1 1 1 0 1 0 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 0 1 1 1 1 1 1 0 0 1 1
F	1 1 1 1 1 0 1 0 1 0 0 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0
L	1 1 1 0 1 1 0 1 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
S	1 1 1 1 1 1 0 1 0 1 1 0 1 1 0 1 1 1 1 1 1 0 1 1 1 1 1 1 1 1 1 1 1 1 1
Z	1 1 1 1 0 1 0 1 0 0 1 0 1 1 0 0 0 1 1 1 0 0 0 0 1 1 1 0 0 0 0 0 1 0 1
	3 3 3 3 3 4 4 4 4 4 4 4 4 4 4 5 5 5 5 5 5 5 5 5 5 6 6 6 6 6 6 6 6 6
	5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8
OG	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
A	1 1 1 1 1 1 0 0 0 0 1 0 1 1 1 1 0 0 0 0 0 1 0 0 1 0 0 1 0 1 0 0 0 1
F	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
L	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
S	1 1 1 1 1 1 1 1 0 1 1 1 1 0 1 1 1 0 1 1 1 1 1 0 1 1 1 1 1 0 1
Z	0 0 0 0 0 0 0 0 0 1 0 1 1 0 0 0 0 0 1 1 0 0 0 1 1 1 1 1 1 1 1 1 1 0

[a] Symbols: OG, Outgroup; A, Australia; F, Africa; L, Laurasia; S, South America; Z, New Zealand.

erty of component data, even though component data see no difference between different trees (Examples 3–5), or between nodes with different redundancy (tables 2, 3). Blind to some aspects of variation in tree structure, component data are perforce blind to some paralogy.

Consider again Example 6 (fig. 1). All five nodes have component data. Node 0 is paralogous. Its component data, like those of all paralogous nodes, are redundant: A + A + A + B + C + C + C, reducing to ABC. Eliminating redundancy, however, does not eliminate paralogy from the component matrix. All comparisons between nodes 1–3 involve node 0 and are paralogous. Without paralogy, there are no data for nodes 1–3. Subtree analysis sees node 0 as paralogous and nodes 1–3 as without data and yields a single subtree, A1(BC1).

Example 6 shows weakness of component data of paralogous nodes. The cladogram includes conflict, variously captured by component and three-item data. Parsimony analysis of both types of matrices yields different trees with less than 100% consistency. For the component matrix, the result, determined by paralogy, differs from that of subtree analysis.

COMBINED MIDGE SUBTREES (FIG. 10)

The 16 subtrees can be combined in one tree (fig. 10), which contains all 19 nonpar-

TABLE 2
Component Data for Nodes (0-68) of Brundin's Combined Cladograms (fig. 8) and Subtrees (fig. 9)[a]

Node	Taxa	Areas	Combined cladograms	Subtrees	Subtree number
0	1–70	AFLSZ+	(AFLSZ)*&	p	—
1	1–47	AFLSZ+	(AFLSZ)*	p	—
2	48–70	AFLSZ+	(AFLSZ)*	p	—
3	1–41	AFLSZ+	(AFLSZ)*	basal	—
4	42–47	AFS+	LZ(AFS)	basal	—
5	48–67	ALSZ+	F(ALSZ)	basal	—
6	68–70	FL+	ALS(FL)	p	—
7	1–37	ASZ+	FL(ASZ)	p	—
8	38–41	FL+	ASZ(FL)	p	—
9	43–47	AS+	FLZ(AS)	p	—
10	48–66	ASZ+	FL(ASZ)	p	—
11	69–70	L+	AFSZ(L)*&	p	—
12	1–25	ASZ+	FL(ASZ)	p	—
13	26–37	ASZ+	FL(ASZ)	p	—
14	39–41	FL+	ASZ(FL)	ASZ(FL)	12
15	43–44	AS	FLZ(AS)	F(AS)	13
16	45–47	AS+	FLZ(AS)	p	—
17	48–65	ASZ+	FL(ASZ)	p	—
18	1–24	ASZ+	FL(ASZ)	p	—
19	26–29	ASZ+	FL(ASZ)	FL(ASZ)	8
20	30–37	AS+	FLZ(AS)	p	—
21	40–41	F+	ALSZ(F)*&	p	—
22	46–47	AS	FLZ(AS)	F(AS)	14
23	48–55	AS+	FLZ(AS)	p	—
24	56–65	SZ+	AFL(SZ)	p	—
25	1–6	ASZ+	FL(ASZ)	FL(ASZ)	1
26	7–24	ASZ+	FL(ASZ)	p	—
27	26–28	AS+	FLZ(AS)	p	—
28	30–32	AS+	FLZ(AS)	p	—
29	33–37	AS+	FLZ(AS)	p	—
30	49–55	AS+	FLZ(AS)	p	—
31	56–58	S+	AFLZ(S)*&	p	—
32	59–65	SZ+	AFL(SZ)	L(SZ)	16
33	1–5	AS+	FLZ(AS)	p	—
34	7–21	ASZ+	FL(ASZ)	p	—
35	22–24	AS+	FLZ(AS)	p	—
36	26–27	AS	FLZ(AS)	FLZ(AS)	8
37	30–31	AS	FLZ(AS)	FL(AS)	9
38	33–34	AS	FLZ(AS)	FL(AS)	10
39	35–37	AS+	FLZ(AS)	p	—
40	49–52	AS+	FLZ(AS)	L(AS)	15
41	53–55	S+	AFLZ(S)*&	p	—
42	56–57	S+	AFLZ(S)*&	p	—
43	59–60	S+	AFLZ(S)*&	p	—
44	61–65	Z+	AFLS(Z)*&	p	—
45	2–5	AS+	FLZ(AS)	p	—
46	7–10	SZ+	AFL(SZ)	p	—
47	11–21	ASZ+	FL(ASZ)	p	—
48	23–24	AS	FLZ(AS)	FL(AS)	7
49	36–37	AS	FLZ(AS)	FL(AS)	11
50	49–50	A+	FLSZ(A)*&	p	—
51	51–52	S+	AFLZ(S)*&	p	—
52	53–54	S+	AFLZ(S)*&	p	—

TABLE 2—(*Continued*)

Node	Taxa	Areas	Combined cladograms	Subtrees	Subtree number
53	61–62	Z+	AFLS(Z)*&	p	—
54	63–65	Z+	AFLS(Z)*&	p	—
55	2–3	S+	AFLZ(S)*&	p	—
56	4–5	AS	FLZ(AS)	FLZ(AS)	1
57	7–8	S+	AFLZ(S)*&	p	—
58	9–10	SZ	AFL(SZ)	FL(SZ)	2
59	11–19	ASZ+	FL(ASZ)	p	—
60	20–21	SZ	AFL(SZ)	FL(SZ)	6
61	63–64	Z+	AFLS(Z)*&	p	—
62	12–19	ASZ+	FL(ASZ)	p	—
63	12–15	SZ+	AFL(SZ)	p	—
64	16–19	ASZ+	FL(ASZ)	FL(ASZ)	5
65	12–13	SZ	AFL(SZ)	FL(SZ)	3
66	14–15	SZ	AFL(SZ)	FL(SZ)	4
67	16–17	Z+	AFLS(Z)*&	p	—
68	18–19	AS	FLZ(AS)	FLZ(AS)	5

[a] Symbols: +, multiple occurrence in one or more area (see table 3); *, node geographically uninformative (component data); &, node uninformative (three-item data); p, paralogous node; and —, node unrepresented in any subtree. Other symbols as in table 1.

alogous nodes—the paralogy-free fraction of the geographic data of the combined cladograms (fig. 8). The combined subtrees (fig. 10) include also the three basal nodes (which thereby become paralogous), as well as one paralogous (and uninformative) node basal to the entire tree. The 19 informative nodes of the subtrees are in distal positions in the combined cladograms (fig. 11), in accord with the expectation that paralogy increases basally in cladograms (see above).

Component data for the nodes of the combined subtrees produce a matrix (table 6) similar to that of table 4 (left). There are four additional characters for nodes 0, 3, 4, and 5, and missing-data entries (question marks) are replaced by 0-entries resulting from paralogous comparisons. Parsimony analysis of the component matrix (table 6) yields two trees (fig. 7, CTComp1 and CTComp2, with strict consensus, Con2). Three-item analysis of the nodes of the combined subtrees (fig. 10) yields 1682 statements. Parsimony analysis of a uniformly weighted matrix yields three trees (fig. 7, CTUW1, CTUW2, CTUW3, with strict consensus, Con3). Parsimony analysis of a fractionally weighted matrix ($\times 10$) yields one tree (fig. 7, CTFW$\times 10$). Results of these analyses of

nodes differ (fig. 7, CTComp1, CTComp2, Con2, CTUW1, CTUW2, CTUW3, Con3, CTFWx10). Such are the ambiguous effects of paralogy, as variously captured by component and three-item data for nodes of combined subtrees. Individual subtrees are paralogy free, but their combination (fig. 10) introduces paralogy at the connecting nodes (nodes 0, 3, 4, 5).

CONCLUSIONS

Some cladogram nodes are geographically paralogous. Parsimony analysis of a matrix with characters for paralogous nodes can yield a tree different from that obtained by parsimony analysis of a matrix for nodes of individual subtrees (the paralogy-free fraction of data). With either type of data (components, three items), the effects of paralogy are unpredictable. We suggest that the paralogy-free fraction contains the only data relevant to area relationship in the cladistic sense. Subtree analysis and parsimony analysis of a matrix for nodes of individual (not combined) subtrees seem the only exact methods presently known to capture and to analyze these data.

TABLE 3
Analysis of Redundancy Eliminated from Component Data for Nodes of Brundin's Combined Cladograms (fig. 8)[a]

| Component Data | Node | Areas | | | | | |
		A	F	L	S	Z	Total
(AFLSZ)	0	11	4	5	34	16	70
	1	9	3	2	23	10	47
	2	2	1	3	11	6	23
	3	7	2	2	20	10	41
LZ(AFS)	4	2	1	0	3	0	6
F(ALSZ)	5	2	0	1	11	6	20
ALS(FL)	6	0	1	2	0	0	3
FL(ASZ)	7	7	0	0	20	10	37
	10	2	0	0	11	6	19
	12	3	0	0	13	9	25
	13	4	0	0	7	1	12
	17	2	0	0	11	5	18
	18	3	0	0	13	8	24
	19	1	0	0	2	1	4
	25	1	0	0	4	1	6
	26	2	0	0	9	7	18
	34	1	0	0	7	7	15
	47	1	0	0	4	6	11
	59	1	0	0	3	5	9
	62	1	0	0	3	4	8
	64	1	0	0	1	2	4
ASZ(FL)	8	0	2	2	0	0	4
	14	0	2	1	0	0	3
FLZ(AS)	9	2	0	0	3	0	5
	15*	1	0	0	1	0	2
	16	1	0	0	2	0	3
	20	3	0	0	5	0	8
	22*	1	0	0	1	0	2
	23	2	0	0	6	0	8
	27	1	0	0	2	0	3
	28	1	0	0	2	0	3
	29	2	0	0	3	0	5
	30	2	0	0	5	0	7
	33	1	0	0	4	0	5
	35	1	0	0	2	0	3
	36*	1	0	0	1	0	2
	37*	1	0	0	1	0	2
	38*	1	0	0	1	0	2
	39	1	0	0	2	0	3
	40	2	0	0	2	0	4
	45	1	0	0	3	0	4
	48*	1	0	0	1	0	2
	49*	1	0	0	1	0	2
	56*	1	0	0	1	0	2
	68*	1	0	0	1	0	2
AFSZ(L)	11	0	0	2	0	0	2
ALSZ(F)	21	0	2	0	0	0	2

TABLE 3—(Continued)

| Component Data | Node | Areas | | | | | |
		A	F	L	S	Z	Total
AFL(SZ)	24	0	0	0	5	5	10
	32	0	0	0	2	5	7
	46	0	0	0	3	1	4
	58*	0	0	0	1	1	2
	60*	0	0	0	1	1	2
	63	0	0	0	2	2	4
	65*	0	0	0	1	1	2
	66*	0	0	0	1	1	2
AFLZ(S)	31	0	0	0	3	0	3
	41	0	0	0	3	0	3
	42	0	0	0	2	0	2
	43	0	0	0	2	0	2
	51	0	0	0	2	0	2
	52	0	0	0	2	0	2
	55	0	0	0	2	0	2
	57	0	0	0	2	0	2
AFLS(Z)	44	0	0	0	0	5	5
	53	0	0	0	0	2	2
	54	0	0	0	0	3	3
	61	0	0	0	0	2	2
	67	0	0	0	0	2	2
FLSZ(A)	50	2	0	0	0	0	2

[a] Each node leads to the indicated number of areas: A (Australia), F (Africa), L (Laurasia), S (South America), and Z (New Zealand). Asterisk indicates nodes with no redundancy: 15, 22, 36–38, 48, 49, 56, 58, 60, 65, 66, and 68.

INTERPRETATION OF NODES

For Brundin, his first and second summaries are different ways to depict the same interpretation (also Brundin, 1974: 294–295). There are two patterns: one older, associated with West Antarctica and New Zealand (fig. 7, P1) and the other younger, associated with East Antarctica and Australia (fig. 7, P2). In Brundin's second summary (fig. 7, S2), node 2 relates New Zealand and West Antarctica and node 4 relates Australia and East Antartica (Brundin, 1966: 452; 1993: 363–364):

the connections between the southern lands have been broken according to a certain sequence beginning with the separation of Southern Africa [node 0]. The next event was the break in the connections between New Zealand and (West) Antarctica [node 2]. The following separation between Tasmania-Australia and (East) Antarctica [node 4] antedates, probably quite considerably, the break between South America and Antarctica.

Subtree 1

|-3-25-56-[4 (A)
            5 (S)
          6 (Z)
          39-40 (FL)

Subtree 2

|-3-58-[9 (Z)
          10 (S)
          39-40 (FL)

Subtree 3

|-3-65-[12 (Z)
          13 (S)
          39-40 (FL)

Subtree 4

|-3-66-[14 (Z)
          15 (S)
          39-40 (FL)

Subtree 5

|-3-64-[16 (Z)
    68-[18 (A)
          19 (S)
          39-40 (FL)

Subtree 6

|-3-60-[20 (S)
          21 (Z)
          39-40 (FL)

Subtree 7

|-3-48-[23 (S)
          24 (A)
          39-40 (FL)

Subtree 8

|-3-19-36-[26 (S)
            27 (A)
          29 (Z)
          39-40 (FL)

Subtree 9

|-3-37-[30 (A)
          31 (S)
          39-40 (FL)

Subtree 10

|-3-38-[33 (A)
          34 (S)
          39-40 (FL)

Subtree 11

|-3-49-[36 (S)
          37 (A)
          39-40 (FL)

Subtree 12

|-3-[1-38 (ASZ)
  14-[39 (L)
        40 (F)

Subtree 13

|-4-[42 (F)
  15-[43 (S)
        44 (A)

Subtree 14

|-4-[42 (F)
  22-[46 (S)
        47 (A)

Subtree 15

|-5-40-[49 (A)
          51 (S)
        67 (L)

Subtree 16

|-5-32-[59 (S)
          61 (Z)
        67 (L)

Fig. 9. Sixteen subtrees derived from Brundin's combined cladograms (fig. 8). If terminal taxon 16 (*P. kiefferi*) is considered the sister of terminal taxon 17, then Subtree 5 alters to (LZ)(AS), which conflicts with other Subtrees; if terminal taxon 16 is considered sister of terminal taxa 17–19, then Subtree 5 alters to L(Z(AS)), which does not conflict (cf. fig 8).

Curiously, Brundin associates no event with node 3, which proves superfluous (no cladistic method sees South America 1 and South America 2 as different). However, Brundin's second summary is an attempt not only to combine both patterns, but also to show a complex pattern of relationship for New Zealand (Brundin, 1970: 46):

> We see from this ["simple diagram," i.e., second summary, but see below] that a group in New Zealand is the sister group of a group occurring in South America [see below; fig. 7, S2A, node 3], or in South America + Australia [see below; fig. 7, S2A, node 4].

In the second summary, there is no node showing "that a group in New Zealand is the sister group of a group occurring in South America" and not in Australia. Presumably Brundin saw node 3 playing this symplesiomorphic role, associating New Zealand and South America 1 (whereas node 3 relates South America 1 and Australia more closely

than to New Zealand). An exact depiction of Brundin's interpretation is more complex (fig. 7, S2A), as suggested by his third summary (fig. 6). No known cladistic method would yield this tree (fig. 7, S2A, or the second sum-

TABLE 4
Component Matrices for Subtrees Derived from Brundin's Combined Cladograms[a]

Nodes of Subtrees	Subtrees
1 1 1 2 2 3 3 3 3 4 4 4 5 5 6 6 6 6 6 4 5 9 2 5 2 6 7 8 0 8 9 6 8 0 4 5 6 8	0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6
OG 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
A 0 1 1 1 1 ? 1 1 1 1 1 1 1 ? ? 1 ? ? 1	2 ? ? ? 2 ? 1 2 1 1 1 0 1 1 1 ?
F 1 0 0 0 ? 0 0 0 ? 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 1 0 0 ? ?
L 1 ? 0 ? 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 1 ? ? 0 0
S 0 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1 1	2 1 1 1 2 1 1 2 1 1 1 0 1 1 1 1
Z 0 ? 1 ? 1 1 0 ? ? ? ? ? 0 1 1 1 1 0	1 1 1 1 1 1 ? 1 ? ? ? 0 ? ? ? 1

[a] Left (two-state characters only), matrix for 19 informative nodes of 16 subtrees (fig. 9). Right (two-state and multistate characters), matrix for 16 subtrees (fig. 9). Symbols as in table 1.

TABLE 5
Three-Item Matrices for Subtrees Derived from Brundin's Combined Cladograms[a]

Nodes of Subtrees

	14	14	14	15	19	19	19	19	22	25	25	25	25	32	36	36	36	37	37	38	38	40
OG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	0	?	?	1	1	1	?	?	1	1	1	?	?	?	1	1	1	1	1	1	1	1
F	1	1	1	0	0	?	0	?	0	0	?	0	?	?	0	?	?	0	?	0	?	?
L	1	1	1	?	?	0	?	0	?	?	0	?	0	0	?	0	?	?	0	?	0	0
S	?	0	?	1	?	?	1	1	1	?	?	1	1	1	1	1	1	1	1	1	1	1
Z	?	?	0	?	1	1	1	1	?	1	1	1	1	1	?	?	0	?	?	?	?	?

	48	48	49	49	56	56	56	58	58	60	60	64	64	64	64	65	65	66	66	68	68	68
OG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	1	1	1	?	?	?	?	1	1	?	?	?	?	?	?	1	1	1
F	0	?	0	?	0	?	?	0	?	0	?	0	?	0	?	0	?	0	?	0	?	?
L	?	0	?	0	?	0	?	?	0	?	0	?	0	?	0	?	0	?	0	?	0	?
S	1	1	1	1	1	1	1	1	1	1	1	?	?	1	1	1	1	1	1	1	1	1
Z	?	?	?	?	?	?	0	1	1	1	1	1	1	1	1	1	1	1	1	?	?	0

Subtrees

	01	01	01	01	01	01	01	02	02	03	03	04	04	05	05	05	05	05	05	05	06	06
OG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	1	?	?	?	?	?	?	?	?	1	1	1	1	1	?	?	?	?
F	0	?	?	0	?	0	?	0	?	0	?	0	?	0	?	?	0	?	0	?	0	?
L	?	0	?	?	0	?	0	?	0	?	0	?	0	?	0	?	?	0	?	0	?	0
S	1	1	1	?	?	1	1	1	1	1	1	1	1	1	1	1	?	?	1	1	1	1
Z	?	?	0	1	1	1	1	1	1	1	1	1	1	?	?	0	1	1	1	1	1	1

	07	07	08	08	08	08	08	08	08	09	09	10	10	11	11	12	12	12	13	14	15	16
OG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	1	1	1	?	?	1	1	1	1	1	1	0	?	?	1	1	1	?
F	0	?	0	?	?	0	?	0	?	0	?	0	?	0	?	1	1	1	0	0	?	?
L	?	0	?	0	?	?	0	?	0	?	0	?	0	?	0	1	1	1	?	?	0	0
S	1	1	1	1	1	?	?	1	1	1	1	1	1	1	1	?	0	?	1	1	1	1
Z	?	?	?	?	0	1	1	1	1	?	?	?	?	?	?	?	?	0	?	?	?	1

[a] Above, matrix for 19 nodes of 16 subtrees (fig. 9). Below, matrix for 16 subtrees (fig. 9). These equivalent matrices are not derived, nor derivable, from the component matrix of table 4 (left), but from the subtrees of figure 9. The matrix for subtrees (below) is derivable also from the component matrix of table 4 (right). Symbols as in table 1.

Fig. 10. Combination of 16 subtrees (fig. 9), including the paralogy-free fraction of Brundin's combined cladograms (fig. 8).

mary) from analysis of Brundin's cladograms (nor is there reason to suggest that he used such to develop his second and third summaries). His cladograms (figs. 2, 3) do not differentiate two sorts of areas for South America and for New Zealand, as indicated by this tree (or two sorts of areas for South America, as indicated in his second summary). Without such differentiation, the best

that could be expected is the combination of the two patterns of the first summary (fig. 7, P1 + 2), or the tree obtained from subtree analysis (fig. 7, Comp). Similar to and confirming Brundin's second summary (less the superfluous node 3 of the second summary), the tree obtained from subtree analysis (fig. 7, Comp) offers the same two nodes relevant to Brundin's interpretation of West Antarctica (node 2) and East Antarctica (node 3).

Brundin (1966: 451) emphasized that for New Zealand midges there are "no direct relationships across the comparatively narrow Tasman Sea." For other groups of organisms such geographic relationship is commonplace. Current models of the geological evolution of New Zealand see the modern condition as derived from collision tectonics mediated by two zones of spreading: an older, extinct Tasman ridge and in the Southern Ocean the currently active ridge separating Antarctica from Australia and New Zealand and continuing northward along the western margin of South America.
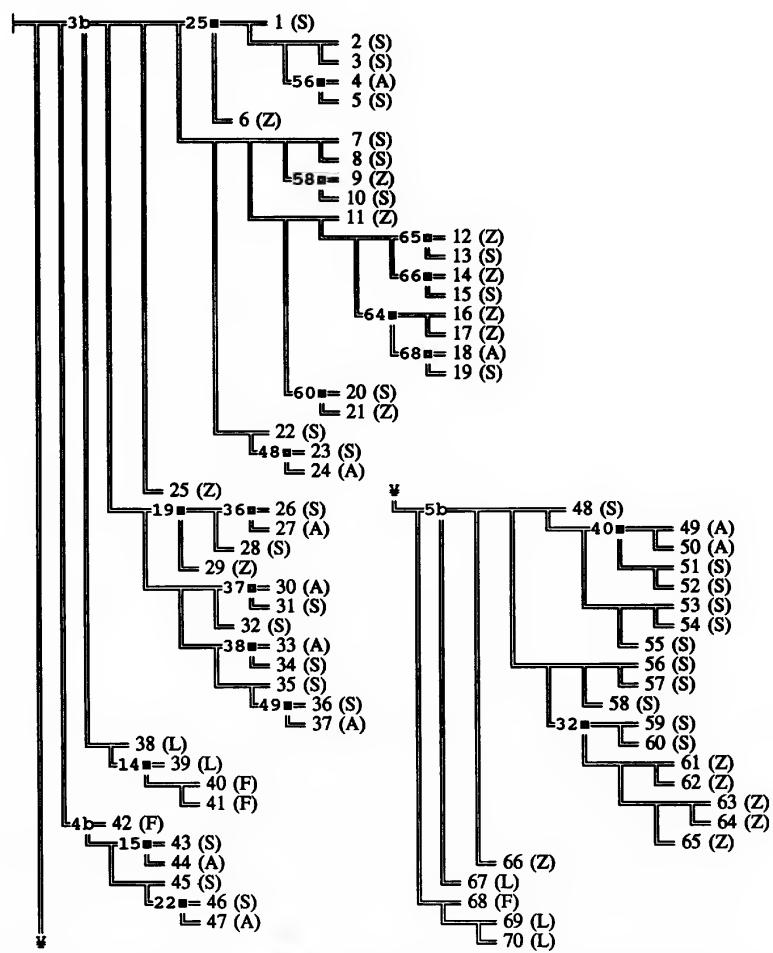
Fig. 11.   Distribution of informative (■) and basal (b) nodes of subtrees (fig. 9) within Brundin's combined cladograms (fig. 8).

An alternative to Brundin's interpretation of the above nodes (2 and 3) is that they relate to the two spreading ridges: node 2 relates to the older Tasman ridge, and node 3 relates to the younger and still active ridge of the Southern Ocean.

For Brundin, southern Africa was isolated early, before differentiation among midges of southern South America, Australia, and New Zealand. Subsequently, New Zealand evidently was isolated early, before differentiation among midges of southern South America and Australia. That there is no trace of "direct relationships" between midges of Australia and New Zealand is no more remarkable than their lack between midges of Africa and South America.

TABLE 6

Component Matrix for Paralogy-Free Fraction of Midge Data[a]

	Nodes of Combined Subtrees																						
	0	0	0	0	1	1	1	2	2	3	3	3	3	4	4	4	5	5	6	6	6	6	6
	0	3	4	5	4	5	9	2	5	2	6	7	8	0	8	9	6	8	0	4	5	6	8
OG	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
A	1	1	1	1	0	1	1	1	1	0	1	1	1	1	1	1	1	0	0	1	0	0	1
F	1	1	1	0	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
L	1	1	0	1	1	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
S	1	1	1	1	0	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1	1
Z	1	1	0	1	0	0	1	0	1	1	0	0	0	0	0	0	0	1	1	1	1	1	0

[a] Matrix for 23 nodes of combined subtrees (fig. 10), containing paralogy-free fraction of midge data. Symbols as in table 1.
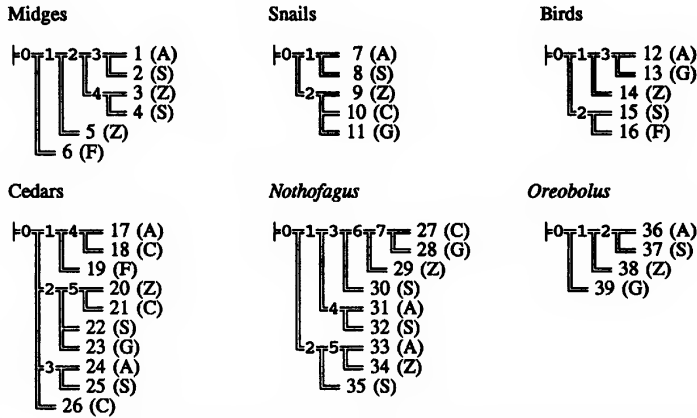
Fig. 12. Cladograms of midges and five other groups of organisms (modified from Craw, 1989: fig. 5): chironomid midges (after Brundin, 1981); land snails of the family Bulimulidae (after Breure, 1979); ratite birds (after Mivart, 1877; Sibley and Ahlquist, 1987); southern cedars of the family Cupressaceae (after Hart, 1987); *Nothofagus* (after Humphries et al., 1986); *Oreobolus* (after Seberg, 1988).

MIDGES AND OTHER ORGANISMS
(fig. 12)

Craw (1989) combined some of Brundin's data and those for five other groups of plants and animals (fig. 12) occurring in the same areas (Australia, New Zealand, South America, Africa) as well as in New Caledonia (C) and New Guinea (G). Treating the six cladograms in effect as individual subtrees, he represented most of their nodes (27 of 29) by component data. With parsimony analysis of the component matrix (table 7, above), he found three trees (fig. 13, C1, C2, C3). He assessed significance through analysis of congruence of the three trees (legend to his fig. 6) with a "geological area cladogram based on the conventional breakup sequence of Gondwana" (fig. 13, CGC). "A set of 999 randomly generated cladograms for the six areas were calculated under both the equiprobable and markovian models and compared for the distances between them and the geological area cladogram" (Craw, 1989: 532). He found that "congruence between these two biological area cladograms [trees C2 and C3 but not tree C1] and the geological tree was just significant at the 5% level" (p. 533).

PARSIMONY ANALYSIS OF NODES

COMBINED CLADOGRAMS (FIG. 14)

The organisms of Craw's six cladograms (fig. 12) relate among themselves (fig. 14) as

animals (node 1), seed plants (node 2), and angiosperms (node 7). Parsimony analysis of a component matrix for nodes (33) of the combined cladograms (table 7, below) yields one tree (fig. 13, Comp). Three-item analysis of the nodes of the combined cladograms (fig. 14) yields 3908 statements. Parsimony analysis of a uniformly weighted matrix yields two trees (fig. 13, UW1 and UW2). The two trees yield a strict consensus (fig. 13, Con2). Parsimony analysis of a fractionally weighted matrix ($\times 10$) yields one tree (fig. 13, FW $\times 10$).

Again, results of parsimony analysis of matrices for nodes of combined cladograms differ (fig. 13, Comp, UW1, UW2, Con2, FW $\times 10$), indicating that data are ambiguous and that the source of ambiguity is geographic paralogy. Within the combined cladograms (fig. 14) some nodes are paralogous (nodes 0, 1, 2, 6, 7, 8, 16, 18, 22). Treating the six cladograms as individual subtrees (without nodes 0, 1, 2, and 7, and with appropriate missing data), Craw eliminates some but not all paralogy (nodes 0, 1, and 2 are uninformative in any case).

INDIVIDUAL SUBTREES (FIG. 15)

Craw's combined cladograms (fig. 14) reduce to 10 subtrees, which together include 15 informative nodes (fig. 15). Parsimony analysis of a component matrix for the 15 nodes of the individual subtrees (table 8, above), yields two trees (fig. 13, C2 and C3),

C1: Africa, N Caledonia, N Guinea, N Zealand, S America, Australia

C2: Africa, N Guinea, N Caledonia, N Zealand, S America, Australia

C3: Africa, N Zealand, N Caledonia, N Guinea, S America, Australia

CGC: Africa, N Caledonia, N Zealand, S America, N Guinea, Australia

Comp: Africa, N Caledonia, N Guinea, N Zealand, S America, Australia

UW1: Africa, N Zealand, S America, N Guinea, N Caledonia, Australia

UW2: Africa, N Zealand, N Guinea, S America, N Caledonia, Australia

Con2: Africa, N Zealand, S America, N Guinea, N Caledonia, Australia

FWx10: Africa, N Zealand, N Guinea, S America, N Caledonia, Australia

STCon: Africa, N Caledonia, N Guinea, N Zealand, S America, Australia

CTComp: Africa, N Caledonia, N Guinea, N Zealand, S America, Australia

CTUW1: N Guinea, N Zealand, S America, Africa, N Caledonia, Australia

CTUW2: N Guinea, N Zealand, S America, N Caledonia, Africa, Australia

CTUW3: N Guinea, N Zealand, S America, Australia, Africa, N Caledonia

Con3: N Guinea, N Zealand, S America, Africa, N Caledonia, Australia

CT FWx10: N Guinea, N Zealand, S America, Africa, N Caledonia, Australia

Fig. 13. Relationships of austral areas, determined by various means. C1, C2 and C3, results (three trees) of parsimony analysis of component matrix (table 7, above) for 27 of 29 nodes of six trees (fig. 12), length 40, ci 67, ri 51 (after Craw, 1989: fig. 6); C2 and C3 are only trees obtained from parsimony analysis of component and three-item matrices (table 8) for 15 nodes of 12 subtrees, with strict consensus of STCon (see below). CGC, geological area cladogram (after Craw, 1989: fig. 6). Comp, result (one tree) of parsimony analysis of component matrix (table 7, below) for 33 nodes of combined cladograms (fig. 14), length 49, ci 67, ri 60. UW1 and UW2, results (two trees) of parsimony analysis of uniformly weighted three-item matrix (4945 statements) for 32 informative nodes of combined cladograms (fig. 14), length 8103, ci 61, ri 36. Con2, strict consensus of two trees (UW1 and UW2). FW × 10, result (one tree) of parsimony analysis of fractionally (× 10) weighted three-item matrix (4945 statements) for 32 informative nodes of combined cladograms (fig. 14), length 69327, ci 61, ri 36. STCon, strict consensus of two trees (same as C2 and C3, above) resulting from parsimony analysis of component matrix (table 8, above) for 15 nodes of 10 subtrees (fig. 15), length 21, ci 71, ri 64; also of uniformly weighted three-item matrix (table 8, below), length 44, ci 77, ri 70; also of fractionally (× 3) weighted 3-item matrix, length 126, ci 76, ri 68. CTComp, result (one tree) of parsimony analysis of component matrix (table 9) for 25 nodes of combined subtrees (fig. 17), length 41, ci 60, ri 51. CTUW1, CTUW2 and CTUW3, results (three trees) of parsimony analysis of uniformly weighted three-item matrix (1671 statements) for 24 informative nodes of combined subtrees (fig. 17), length 2761, ci 61, ri 37. Con3, strict consensus of three trees (CTUW1, CTUW2, CTUW3). CTFW × 10, result (one tree) of parsimony analysis of fractionally (× 10) weighted three-item matrix (1671 statements) for 24 informative nodes of combined subtrees (fig. 17), length 22330, ci 62, ri 38.

with strict consensus grouping New Zealand, New Caledonia, and New Guinea (fig. 13, StCon). Parsimony analysis of uniformly and fractionally weighted three-item matrices for nodes of the 10 individual subtrees (table 8, below) yields the same result (fig. 13, C2 and C3). As in the case of Brundin's combined cladograms, informative nodes of subtrees are in terminal positions in the 33-node combination (fig. 16).

COMBINED SUBTREES (FIG. 17)

The 10 subtrees combine in one tree (fig. 17) somewhat simpler (25 versus 33 nodes) than the combination of the original six cladograms (fig. 14). Not enough paralogy is eliminated from this simpler tree, however, for parsimony analysis of a component matrix for its nodes (table 9) to yield a result (fig. 13, CTComp) different from that obtained above from parsimony analysis of a component matrix for the nodes of the 33-node combination of six cladograms (fig. 13, Comp). Three-item analysis of the nodes of the combined subtrees (fig. 17) yields 1671 statements. Parsimony analysis of a uniformly weighted matrix yields three trees (fig. 13, CTUW1, CTUW2, CTUW3), with strict consensus grouping Australia, New Caledo-
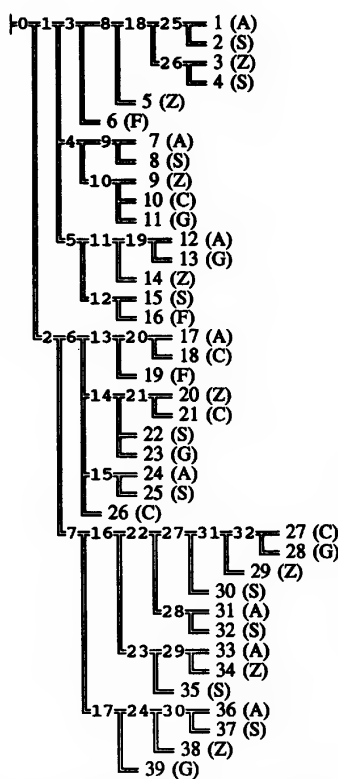
Fig. 14. Combination of Craw's (1989) six cladograms of midges and other organisms (fig. 12).
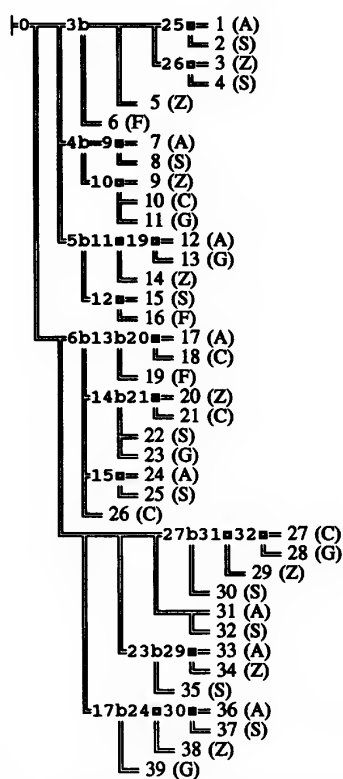
Fig. 16. Distribution of informative (■) and basal (b) nodes of subtrees (fig. 15) within combination of six cladograms of midges and other organisms (fig. 14).
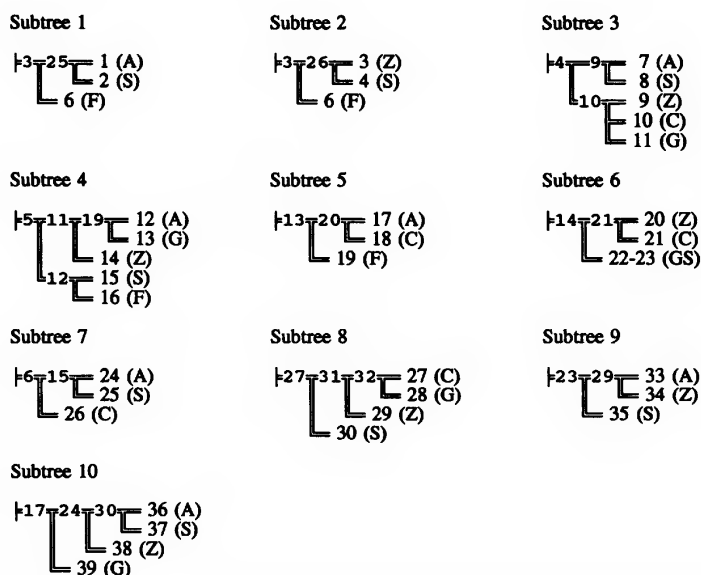
Fig. 15. Ten subtrees derived from combined cladograms of midges and other organisms (fig. 14).

TABLE 7
Component Matrices for Craw's (1989) Cladograms[a]

	Midges	Snails	Birds	Cedars	Nothofagus	Oreobolus
	Nodes of Cladograms					
	0234	012	0123	012345	1234567	012
OG	0000	000	0000	000000	0000000	000
A	1110	110	1101	110110	1101100	111
F	1000	???	1010	110000	???????	???
S	1111	110	1010	101100	1111000	111
Z	1101	101	1100	101001	1110110	110
C	????	101	????	111011	1010011	???
G	????	101	1101	101000	1010010	100

Nodes of Combined Cladograms

```
   0000000000011111111111122222222222333
   0123456789012345678901234567890123456789012
OG 000000000000000000000000000000000
A  111111111110101011111101111001100
F  111101100000110000000000000000000
S  111111111100101111100011111110100
Z  111111111011010110111001111011010
C  111010110010011010001110000100011
G  111011110011001011010010000100011
```

[a] Above, matrix for 27 nodes of Craw's six cladograms (fig. 12, modified from Craw, 1989: table 4). Below, matrix for 33 nodes of Craw's combined cladograms (fig. 14). Craw's matrix (1989: table 4) omits data for two nodes (node 1, Midges cladogram; node 0, *Nothofagus* cladogram). Symbols: C, New Caledonia; G, New Guinea. Other symbols as in table 1.

TABLE 8
Component and Three-Item Matrices for Subtrees Derived from Craw's Cladograms[a]

Nodes of Subtrees

```
   0 0 00 000 0 0 0 00 0 11
   1 2 33 444 5 6 7 88 9 00
OG 0 0 00 000 0 0 0 00 0 00
A  1 ? 10 110 1 ? 1 ?? 1 11
F  0 0 ?? 001 0 ? ? ?? ? ??
S  1 1 10 001 ? 0 1 00 0 11
Z  ? 1 01 100 ? 1 ? 10 1 10
C  ? ? 01 ??? 1 1 0 11 ? ??
G  ? ? 01 110 ? 0 ? 11 ? 00
```

Subtrees

```
   0 0 000000000 0000000
   1 2 333333333 4444444
OG 0 0 000000000 0000000
A  1 ? 111?0?0?0 11??111
F  0 0 ????????? 0?0?0??
S  1 1 111?0?0?0 ?0?0?0?
Z  ? 1 0??111??? 1111??0
C  ? ? ?0?11??11 ???????
G  ? ? ??0??1111 ??11111

   000 0 00 0 0000 0 1111
   444 5 66 7 8888 9 0000
OG 000 0 00 0 0000 0 0000
A  0?? 1 ?? 1 ???? 1 1?11
F  111 0 ?? ? ???? ? ????
S  111 ? 0? 1 000? 0 ?111
Z  ?0? ? 11 ? 11?0 1 110?
C  ??? 1 11 0 1?11 ? ????
G  ??0 ? ?0 ? ?111 ? 00?0
```

[a] Above, component matrix for 15 informative nodes of 10 subtrees (fig. 15). Below, three-item matrix for 10 subtrees (fig. 15). Symbols as in tables 1 and 7.

nia, and Africa (fig. 13, Con3). Parsimony analysis of a fractionally weighted matrix ($\times 10$) yields one tree (fig. 13, CTFWx10). Results of these analyses differ (fig. 13, CTComp, CTUW1, CTUW2, CTUW3, Con3, CTFW $\times$ 10). Such, again, are the ambiguous effects of paralogy, as variously captured by component and three-item data for the nodes of the combined subtrees.

SIGNIFICANCE OF OTHER ORGANISMS

With addition of groups of organisms occurring in New Caledonia and New Guinea, results of parsimony analysis of the enlarged matrix for nodes of subtrees do not contradict the results of analysis of the matrix for the nodes of subtrees of midges, but indicate merely that the added areas (New Caledonia and New Guinea) relate to New Zealand. For Craw (1989: 533), however, his results supply evidence for "three different views of the biogeographic classification of New Zealand in relation to Australia and South America" (the other two views are based on analyses not of nodes of cladograms but of presence/absence data of genera and families of plants and animals).

Unlike the midge data, nevertheless, there is conflict in the data added, not only regarding relationships of the added areas and New Zealand but also regarding relationships of New Zealand and South America on the one hand and relationships of Australia and New Guinea on the other (fig. 12): ratite birds relate South America and Africa more closely than to New Zealand; cedars relate New Zealand and New Caledonia more closely than to New Guinea, whereas *Nothofagus* relates New Guinea and New Caledonia more closely than to New Zealand, and Australia and
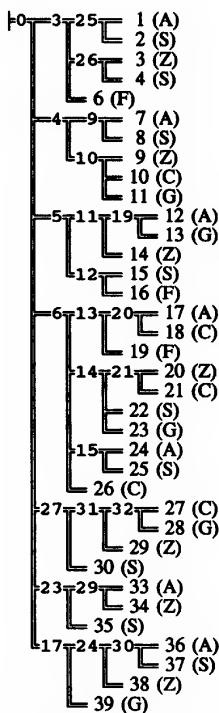
Fig. 17. Combination of 10 subtrees (fig. 15) that includes the paralogy-free fraction of combined cladograms of midges and other organisms (fig. 14).

Fig. 18. Combination of cladograms for seven species groups of North American fishes (after Mayden, 1988). Cladograms (nodes): 4, *Nocomis biguttatus* group; 5, *Fundulus catenatus* group; 7, *Notropis leuciodus* group; 8, *Luxilus* zonatus group; 11, subgenus *Imostoma* of Percina; 17, *Etheostoma variatum* group; 18, subgenus *Ozarkia* of *Etheostoma*. Other nodes: 0, Teleostei; 1, Cyprinidae; 2, Neoteleostei; 3, *Notropis* (old usage); 6, Percidae; 12, *Etheostoma*.

New Zealand more closely than to South America; *Oreobolus* relates New Zealand more closely to Australia and South America than to New Guinea. Such conflict, evident

TABLE 9
Component Matrix for Paralogy-Free Fraction of Craw's Data[a]

```
                  Nodes of Combined Subtrees
     0 0 0 0 0 0 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 3 3 3
     0 3 4 5 6 9 0 1 2 3 4 5 7 9 0 1 3 4 5 6 7 9 0 1 2

OG   0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
A    1 1 1 1 1 1 1 0 1 0 1 0 1 1 1 1 0 1 1 1 0 0 1 1 0
F    1 1 0 1 1 0 0 0 1 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
S    1 1 1 1 1 1 0 0 1 0 1 1 1 0 0 0 1 1 1 1 1 0 0 1 0
Z    1 1 1 1 1 0 1 1 0 0 1 0 1 0 0 1 1 1 0 1 1 1 0 1 0
C    1 0 1 0 1 0 1 0 0 1 1 0 0 0 1 1 0 0 0 0 1 0 0 1 1
G    1 0 1 1 1 0 1 1 0 0 1 0 1 1 0 0 0 0 0 0 1 0 0 1 1
```

[a] Matrix for 25 nodes of combined subtrees, containing paralogy-free fraction of data for midges and other organisms (fig. 17). Symbols as in tables 1 and 7.

without parsimony analysis of any matrix, accords with Craw's views about the composite nature of New Zealand. The amount of conflict is reflected in the low consistency indices (74–77) of the various analyses of matrices for nodes of subtrees that represent the paralogy-free fraction of the data.

NORTH AMERICAN FRESHWATER FISHES
(Mayden, 1988)

PARSIMONY ANALYSIS OF NODES

COMBINED FISH CLADOGRAMS
(FIGS. 18, 19 LEFT)

In the most ambitious undertaking of its kind, Mayden (1988) offered a geographic analysis of cladograms for seven species groups of fishes distributed among 34 major rivers of the central United States. Cladograms for the seven species groups can be
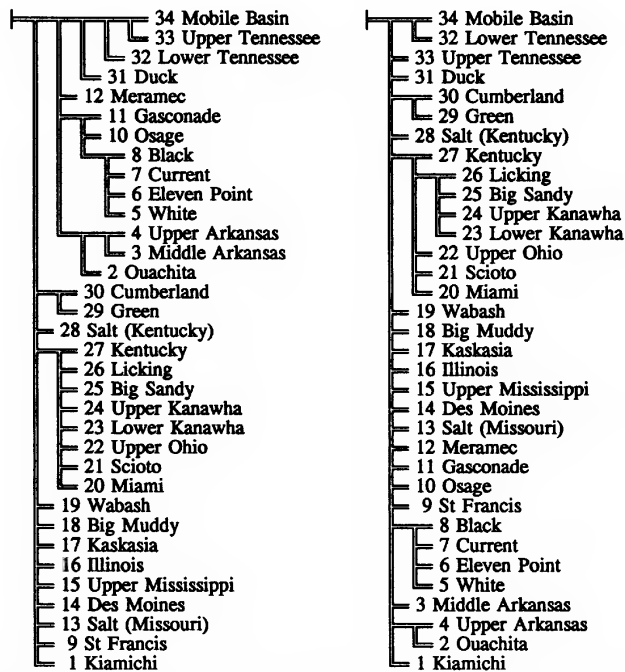
Fig. 19. Strict consensus trees. **Left** (10 nodes), of 89 trees (length 72, ci 38, ri 82), from parsimony analysis of component matrix (table 10, no assumption 2) for 28 nodes of combined cladograms (fig. 18). **Right** (6 nodes), of 58 trees (length 75, ci 37, ri 80) from parsimony analysis of component matrix (table 10, assumption 2) for 28 nodes of combined cladograms (fig. 22).

combined as one cladogram of 29 terminal taxa and 28 nodes (fig. 18). Of these, seven nodes are paralogous (nodes 0–3, 6, 12, 24). Table 10 is a component matrix for all 28 nodes. Parsimony analysis of this matrix (no assumption 2) yields 89 trees, of which the strict consensus has 10 informative nodes (fig. 19, left).

SUBTREE ANALYSIS OF COMBINED FISH CLADOGRAMS (FIG. 20)

Mayden's combined cladograms (fig. 18) reduce to eight subtrees (fig. 20), which together include 15 informative nodes. Subtrees 1–6 are no different from the cladograms of six of Mayden's seven species groups. Subtrees 7 and 8 represent the paralogy-free fractions of the cladogram for the remaining (seventh) species group, which contains one geographically paralogous node (fig. 18, node 24)—the only paralogous node in Mayden's seven cladograms, if each of them is considered separately from the others (as an individual subtree).

INDIVIDUAL FISH SUBTREES (FIG. 21)

For the 15 nodes of the individual subtrees (fig. 20), parsimony analysis of a component matrix (table 11) yields 5302+ (overflow) trees (length 16), of which the strict consensus contains two informative nodes (areas 1–33; areas 20–27). Among the 5302 trees, the degree of resolution (number of informative nodes) ranges from 7 to 24 (table 12). One tree (fig. 21, left) is least resolved (seven nodes). This tree has one spurious node, and some areas are placed higher in the tree than is warranted by the data. This tree (length 16) collapses to yield a tree of least resolution, also of length 16 (fig. 21, right). This minimal tree (six nodes) is offered as an exact result of parsimony analysis of the matrix for the paralogy-free fraction of data of all nodes of the seven cladograms of Mayden, as rendered by subtree analysis (fig. 20 and table 11). This minimal tree is nearly the same as that previously obtained by hand resolution for assumption 2 (Nelson and Ladiges, 1991a: fig. 3B; see below).

TABLE 10
Component Matrix for Mayden's (1988) Combined Cladograms[a]

	0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2
Nodes of Combined Cladograms:	0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7
OUTGROUP	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
01 KIAMICHI	1 1 1 1 1 1 0 1 1 1 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0
02 OUACHITA	1 1 1 1 1 1 1 1 0 1 1 1 1 1 0 0 1 0 1 0 0 1 0 0 1 0 1 0
03 MIDDLE ARKANSAS	1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 0 0 1 1 0 1 0 0 1 1 1
04 UPPER ARKANSAS	1 1 1 1 1 0 1 1 1 1 0 0 1 1 0 1 0 0 1 1 0 1 0 0 1 0 1 0
05 WHITE	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 1 1 1 0 1 1 1 1 0 1
06 ELEVEN POINT	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 0 1 1 1 1 0 1
07 CURRENT	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 0 1 1 1 1 0 1
08 BLACK	1 1 1 1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 1 1 0 1 1 1 1 0 1
09 ST FRANCIS	1 1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 1 0 0 0 0 0 0 0 0 0 0 0
10 OSAGE	1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 1 0 1 1 0 1 1 0 1 1 1 0
11 GASCONADE	1 1 1 1 1 1 1 1 1 1 1 0 1 0 1 0 1 1 0 1 1 0 1 1 0 1 1 0
12 MERAMEC	1 1 1 1 1 1 1 1 1 1 1 0 1 1 0 1 0 1 0 0 1 0 0 1 0 0 0 0
13 SALT (MISSOURI)	1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 DES MOINES	1 1 0 1 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
15 UPPER MISSISSIPPI	1 1 0 1 1 0 0 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0
16 ILLINOIS	1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
17 KASKASIA	1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
18 BIG MUDDY	1 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
19 WABASH	1 1 1 0 1 1 1 0 0 1 1 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0
20 MIAMI	1 1 1 0 1 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0
21 SCIOTO	1 1 1 0 1 0 1 0 0 1 0 0 1 0 0 1 0 0 0 1 0 1 0 0 0 0 0 0
22 UPPER OHIO	1 1 1 0 1 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0
23 LOWER KANAWHA	1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0
24 UPPER KANAWHA	1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0
25 BIG SANDY	1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0
26 LICKING	1 0 1 0 0 0 1 0 0 0 0 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0
27 KENTUCKY	1 1 1 0 1 1 1 0 0 1 1 0 1 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0
28 SALT (KENTUCKY)	1 0 1 0 0 1 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
29 GREEN	1 1 1 1 1 1 0 1 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0
30 CUMBERLAND	1 1 1 1 1 1 0 1 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0
31 DUCK	1 1 1 1 1 1 1 1 0 0 1 0 1 1 0 0 0 1 1 0 0 1 0 0 0 0 0 0
32 LOWER TENNESSEE	1 1 1 1 0 1 1 1 1 0 1 0 1 1 1 0 0 1 1 0 0 1 0 0 0 0 0 0
33 UPPER TENNESSEE	1 1 1 1 0 1 1 1 1 0 1 1 1 1 1 0 1 1 0 0 0 0 0 0 0 0 0 0
34 MOBILE BASIN	1 1 1 1 0 1 1 1 1 0 0 1 1 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0

[a] Matrix for 28 nodes of combined cladograms for seven species groups of North American freshwater fishes (fig. 18), no assumption 2. For assumption 2 (fig. 22), certain 1-entries become 0-entries: area 03, node 26; area 04, nodes 8, 15, 19; area 32, nodes 5, 10, 13, 18, 21; area 33, nodes 8, 14.

Three-item analysis of the 15 nodes of individual subtrees yields 971 statements. Parsimony analysis of a uniformly weighted matrix yields 1757+ (overflow) trees, length 993, ci 97, ri 97, of which the strict consensus has one informative node (areas 1–33). Of the 1757 trees, least resolved are 14 trees (24 nodes, range 24–32). Of the 14 trees, one is least informative (2844 independent statements, range 2844–3068). This tree collapses to the same minimal tree, length 993 (fig. 21, right). Parsimony analysis of a fractionally weighted ($\times 10$) matrix yields 1758+ (overflow) trees, length 4620, ci 95, ri 95, of which the strict consensus has one informative node (areas 1–33). Of the 1758 trees, one is least resolved (22 nodes, range 22–32). This tree collapses to the same minimal tree, length 4620 (fig. 21, right).

RATIONALE OF ASSUMPTION 2

As noted previously (Nelson and Ladiges, 1991a: 48), conflicting elements among the seven cladograms are removed by applica-

tion of assumption 2, which reduces the geographic distribution of six species (fig. 22: species 2, 5, 8, 13, 24, 28).

SUBGENUS *OZARKIA* OF *ETHEOSTOMA* (FIGS. 18, 22: NODE 18)

There is conflict with respect to area 3 (fig. 18): node 27 relates areas 3 and 5–8; node 26 relates areas 3–4 and 2. For node 27, *E. punctulatum* A (species 27) is endemic in area 3; for node 26, *E. cragini* (species 28) is widespread in areas 3–4. Under assumption 2, *E. cragini* is seen as possibly endemic to area 4, with secondary dispersal (mobilism) into area 3. Application of assumption 2 reduces the distribution of *E. cragini* to area 4 (the relationship of area 3 is determined by the endemic, *E. punctulatum* A, of which the distribution is unreducible).

LUXILUS ZONATUS SPECIES-GROUP (FIGS. 18, 22: NODE 8)

There is conflict with respect to area 4 (fig. 18): node 19 relates areas 1, 3–4 and 5; node 26 (fig. 22) relates areas 2 and 4 (note that area 3 is removed from the distribution of *E. cragini*, see above). For node 19, *L. pilsbryi* (species 7) is endemic in area 5, and *L. cardinalis* (species 8) is widespread in areas 1, 3–4; for node 26, *E. cragini* (species 28) is possibly endemic in area 4 (see above), and *E. pallididorsum* (species 29) is endemic in area 2. Under assumption 2, *L. cardinalis* is seen as possibly endemic to areas 1 and 3, with secondary dispersal (mobilism) into area 4. Application of assumption 2 reduces the distribution of *L. cardinalis* to areas 1 and 3 (the relationship of area 4 is determined by the possible endemic, *E. cragini*, of which the distribution is not further reducible).

There is conflict with respect to area 33 (fig. 18): node 14 relates areas 32–33 and 34; node 16 relates areas 2, 5–9, 19, and 33. For node 14, *L. zonistius* (species 4) is endemic in area 34, and *L. coccogenis* (species 5) is widespread in areas 32–33; for node 16, *P. tanasi* (species 16) is endemic in area 33. Under assumption 2, *L. coccogenis* is seen as possibly endemic in area 32, with secondary dispersal (mobilism) into area 33. Application of assumption 2 reduces the distribution of *L. coccogenis* to area 32 (the relationship of area 33 is determined by the endemic, *P.*
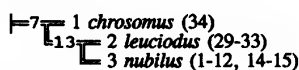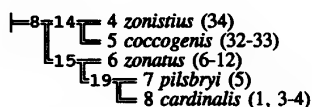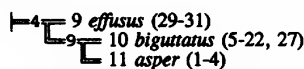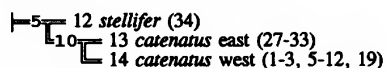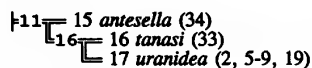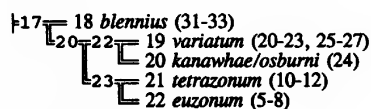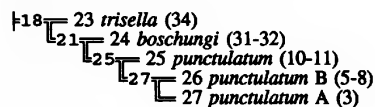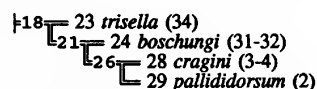
Fig. 20. Subtrees derived from combined cladograms (fig. 18, no assumption 2).

tanasi, of which the distribution is unreducible).

OTHER OCCURRENCES IN AREA 32

Under assumption 2, other widespread taxa occurring in area 32 are seen as possibly occurring there because of secondary dispersal (mobilism): *N. leuciodus* (fig. 18: node 13, species 2), *F. catenatus* east (node 10, species 13), and *E. boschungi* (node 21, species 24). Application of assumption 2 reduces the distribution of each of these species (fig. 22) such that area 32 is removed therefrom (the relationship of area 32 is determined by the
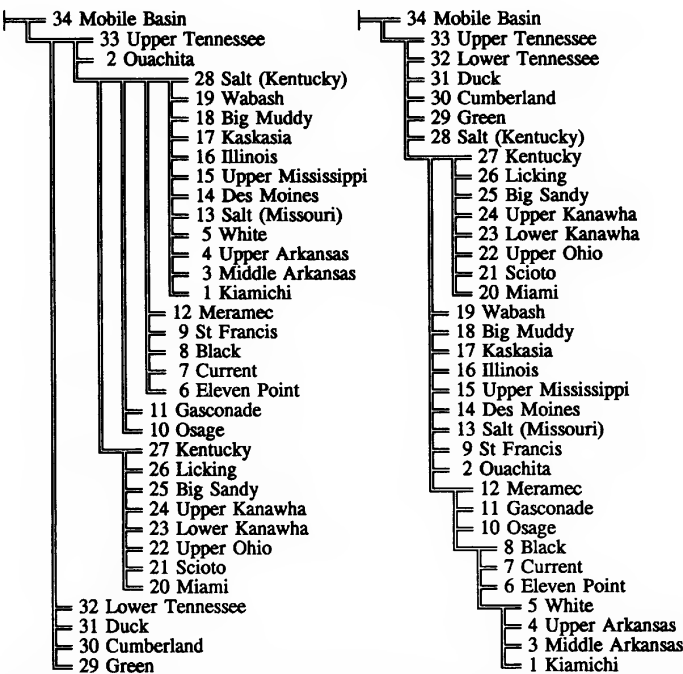
Fig. 21. Results of parsimony analysis of component matrix (table 11, no assumption 2) for 15 nodes of subtrees (fig. 20). **Left** (seven nodes), least resolved (table 12) of 5302+ (overflow) trees (length 16, ci 93, ri 97). **Right** (six nodes), minimal tree (length 16) derived from least resolved (left) tree by collapse of nodes and more basal placement of some areas. The same minimal tree results from parsimony analysis of three-item matrices (see text).

possible endemic, *L. coccogenis* of node 14, of which the distribution is not further reducible).

PARSIMONY ANALYSIS OF NODES, ASSUMPTION 2

COMBINED FISH CLADOGRAMS (FIGS. 22, 19 RIGHT)

For the combined cladograms (fig. 18, 22), application of assumption 2 eliminates paralogy for one node (node 24), leaving six paralogous nodes (nodes 0–3, 6, 12). Parsimony analysis of a component matrix for the nodes of the combined cladograms (table 10, assumption 2) yields 58 trees, of which the strict consensus has six informative nodes (fig. 19, right). Comparison of figure 19 left and right shows the effects of application of assumption 2 on the results of parsimony analysis of a component matrix for all nodes of the combined cladograms:

1. overall reduction in resolution (from 10 to 6 informative nodes);

2. increased resolution within one grouping (areas 20–27) with addition of 1 novel node (areas 23–26);
3. different groupings of the same areas (areas 2 and 4; areas 32 and 34).

INDIVIDUAL FISH SUBTREES (FIGS. 23, 24)

For the 15 nodes of the individual subtrees, some affected by assumption 2 (fig. 23), a component matrix (table 13) yields 5299+ (overflow) trees (length 15), of which the strict consensus has two informative nodes (areas 32, 34; areas 1–31, 33). Among the 5299 trees, the degree of resolution (number of informative nodes) ranges from 9 to 28 (table 14). Four trees are least resolved (nine nodes). The strict consensus of these four trees has three informative nodes (areas 32, 34; areas 29–31, 33; areas 1–28). Among these four trees, the amount of their data (number of independent three-item statements) ranges from 1340 to 1508 (table 15). One tree (fig. 24, left) is least informative (1340 state-

TABLE 11

Component Matrix for Subtrees Derived from Mayden's Combined Cladograms[a]

Nodes of Subtrees:	0 1 1 1 1 1 1 1 2 2 2 2 2 2 2 9 0 3 4 5 6 9 0 1 1 2 3 5 6 7
OUTGROUP	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
01 KIAMICHI	1 1 1 0 1 ? 1 ? ? ? ? ? ? ? ?
02 OUACHITA	1 1 1 ? ? 1 ? ? ? 1 ? ? ? 1 ?
03 MIDDLE ARKANSAS	1 1 1 0 1 ? 1 ? 1 1 ? ? 1 1 1
04 UPPER ARKANSAS	1 ? 1 0 1 ? 1 ? ? 1 ? ? ? 1 ?
05 WHITE	1 1 1 0 1 1 1 1 1 ? 0 1 1 ? 1
06 ELEVEN POINT	1 1 1 0 1 1 0 1 1 ? 0 1 1 ? 1
07 CURRENT	1 1 1 0 1 1 0 1 1 ? 0 1 1 ? 1
08 BLACK	1 1 1 0 1 1 0 1 1 ? 0 1 1 ? 1
09 ST FRANCIS	1 1 1 0 1 1 0 ? ? ? ? ? ? ? ?
10 OSAGE	1 1 1 0 1 ? 0 1 1 ? 0 1 1 ? 0
11 GASCONADE	1 1 1 0 1 ? 0 1 1 ? 0 1 1 ? 0
12 MERAMEC	1 1 1 0 1 ? 0 1 ? ? 0 1 ? ? ?
13 SALT (MISSOURI)	1 ? ? ? ? ? ? ? ? ? ? ? ? ? ?
14 DES MOINES	1 ? 1 ? ? ? ? ? ? ? ? ? ? ? ?
15 UPPER MISSISSIPPI	1 ? 1 ? ? ? ? ? ? ? ? ? ? ? ?
16 ILLINOIS	1 ? ? ? ? ? ? ? ? ? ? ? ? ? ?
17 KASKASIA	1 ? 1 ? ? ? ? ? ? ? ? ? ? ? ?
18 BIG MUDDY	1 ? ? ? ? ? ? ? ? ? ? ? ? ? ?
19 WABASH	1 1 ? ? ? 1 ? ? ? ? ? ? ? ? ?
20 MIAMI	1 ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
21 SCIOTO	1 ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
22 UPPER OHIO	1 ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
23 LOWER KANAWHA	? ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
24 UPPER KANAWHA	? ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
25 BIG SANDY	? ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
26 LICKING	? ? ? ? ? ? ? 1 ? ? 1 0 ? ? ?
27 KENTUCKY	1 1 ? ? ? ? ? 1 ? ? 1 0 ? ? ?
28 SALT (KENTUCKY)	? 1 ? ? ? ? ? ? ? ? ? ? ? ? ?
29 GREEN	0 1 1 ? ? ? ? ? ? ? ? ? ? ? ?
30 CUMBERLAND	0 1 1 ? ? ? ? ? ? ? ? ? ? ? ?
31 DUCK	0 1 1 ? ? ? ? 0 1 1 0 0 0 0 0
32 LOWER TENNESSEE	? 1 1 1 0 ? 0 0 1 1 0 0 0 0 0
33 UPPER TENNESSEE	? 1 1 1 0 1 0 0 ? ? 0 0 ? ? ?
34 MOBILE BASIN	? 0 0 1 0 0 0 ? 0 0 ? ? 0 0 0

[a] Matrix for 15 informative nodes of eight subtrees (fig. 20), no assumption 2.

TABLE 12

Number of Nodes of 5302 Trees Derived from Mayden's Data[a]

Number of		Number of	
Nodes	Trees	Nodes	Trees
24	441	15	677
23	510	14	323
22	495	13	179
21	458	12	116
20	375	11	55
19	365	10	48
18	325	9	39
17	290	8	11
16	594	7	1

[a] Resulting from parsimony analysis of component matrix for subtree nodes (table 11).

tion 2 (fig. 23), yields 913 statements. Parsimony analysis of a uniformly weighted matrix yields 1757+ (overflow) trees, length 913, ci 100, ri 100, of which the strict consensus has two informative nodes (areas 1–31, 33; areas 32 and 34). Of the 1757 trees, one tree is least resolved (16 nodes, range 16–32). This tree collapses to the same minimal tree, length 913 (fig. 24, right). Parsimony analysis of a fractionally weighted (×10) matrix yields

Fig. 22. Combination of cladograms for seven species groups (cf. fig. 18). Asterisk (*), species with geographic distribution reduced according to assumption 2.

ments). This tree (length 15) collapses to yield a tree of least resolution, also of length 15 (fig. 24, right). This minimal tree (eight nodes) is offered as an exact result of parsimony analysis of the matrix for the paralogy-free fraction of data of all nodes of the seven cladograms of Mayden, as rendered by subtree analysis for assumption 2. This minimal tree is the same as that previously obtained by hand resolution (Nelson and Ladiges, 1991a: fig. 3B).

Three-item analysis of the 15 nodes of individual subtrees, some affected by assump-

Subtree 1

```
|—7┌— 1 chrosomus (34)
   └13┌— 2 leuciodus (29-31, 33)*
      └— 3 nubilus (1-12, 14-15)
```

Subtree 2

```
|—8┌—14┌— 4 zonistius (34)
   |    └— 5 coccogenis (32)*
   └15┌— 6 zonatus (6-12)
      └19┌— 7 pilsbryi (5)
         └— 8 cardinalis (1, 3)*
```

Subtree 3

```
|—4┌— 9 effusus (29-31)
   └9┌— 10 biguttatus (5-22, 27)
     └— 11 asper (1-4)
```

Subtree 4

```
|—5┌— 12 stellifer (34)
   └10┌— 13 catenatus east (27-31, 33)*
      └— 14 catenatus west (1-3, 5-12, 19)
```

Subtree 5

```
|—11┌— 15 antesella (34)
    └16┌— 16 tanasi (33)
       └— 17 uranidea (2, 5-9, 19)
```

Subtree 6

```
|—17┌— 18 blennius (31-33)
    └20┌—22┌— 19 variatum (20-23, 25-27)
       |   └— 20 kanawhae/osburni (24)
       └23┌— 21 tetrazonum (10-12)
          └— 22 euzonum (5-8)
```

Subtree 7

```
|—18┌— 23 trisella (34)
    └21┌— 24 boschungi (31)*
       └24┌—25┌— 25 punctulatum (10-11)
          |   └27┌— 26 punctulatum B (5-8)
          |      └— 27 punctulatum A (3)
          └26┌— 28 cragini (4)*
             └— 29 pallididorsum (2)
```

Fig. 23. Subtrees derived from combined cladograms (fig. 22). Asterisk (*), species with geographic distribution reduced according to assumption 2 (cf. fig. 22).

1757+ (overflow) trees, length 4206, ci 100, ri 100, of which the strict consensus has two informative nodes (areas 1–31, 33; areas 32 and 34). Of the 1757 trees, one tree is least resolved (19 nodes, range 19–32). This tree collapses to the same minimal tree, length 4206 (fig. 24, right).

Comparison of figures 21 (right) and 24 (right) shows effects of application of assumption 2 on results of parsimony analysis of matrices (component and three-item) for nodes of individual subtrees:

1. overall increase in resolution (from six to eight informative nodes);

TABLE 13
Component Matrix for Subtrees Derived from Mayden's Combined Cladograms[a]

Nodes of Subtrees:	09	10	13	14	15	16	19	20	21	22	23	24	25	26	27
OUTGROUP	0	0	0	0	0	0	0	0	0	0	0	0	0	0	0
01 KIAMICHI	1	1	1	0	1	?	1	?	?	?	?	?	?	?	?
02 OUACHITA	1	1	1	?	?	1	?	?	1	?	?	1	0	1	0
03 MIDDLE ARKANSAS	1	1	1	0	1	?	1	?	1	?	?	1	1	0	1
04 UPPER ARKANSAS	1	?	1	?	?	?	?	?	1	?	?	1	0	1	0
05 WHITE	1	1	1	0	1	1	1	1	1	0	1	1	1	0	1
06 ELEVEN POINT	1	1	1	0	1	1	0	1	1	0	1	1	1	0	1
07 CURRENT	1	1	1	0	1	1	0	1	1	0	1	1	1	0	1
08 BLACK	1	1	1	0	1	1	0	1	1	0	1	1	1	0	1
09 ST FRANCIS	1	1	1	0	1	1	0	?	?	?	?	?	?	?	?
10 OSAGE	1	1	1	0	1	?	0	1	1	0	1	1	1	0	0
11 GASCONADE	1	1	1	0	1	?	0	1	1	0	1	1	1	0	0
12 MERAMEC	1	1	1	0	1	?	0	1	?	0	1	?	?	?	?
13 SALT (MISSOURI)	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?
14 DES MOINES	1	?	1	?	?	?	?	?	?	?	?	?	?	?	?
15 UPPER MISSISSIPPI	1	?	1	?	?	?	?	?	?	?	?	?	?	?	?
16 ILLINOIS	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?
17 KASKASIA	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?
18 BIG MUDDY	1	?	?	?	?	?	?	?	?	?	?	?	?	?	?
19 WABASH	1	1	?	?	?	1	?	?	?	?	?	?	?	?	?
20 MIAMI	1	?	?	?	?	?	?	1	?	1	0	?	?	?	?
21 SCIOTO	1	?	?	?	?	?	?	1	?	1	0	?	?	?	?
22 UPPER OHIO	1	?	?	?	?	?	?	1	?	1	0	?	?	?	?
23 LOWER KANAWHA	?	?	?	?	?	?	?	1	?	1	0	?	?	?	?
24 UPPER KANAWHA	?	?	?	?	?	?	?	1	?	1	0	?	?	?	?
25 BIG SANDY	?	?	?	?	?	?	?	1	?	1	0	?	?	?	?
26 LICKING	?	?	?	?	?	?	?	1	?	1	0	?	?	?	?
27 KENTUCKY	1	1	?	?	?	?	1	1	?	1	0	?	?	?	?
28 SALT (KENTUCKY)	?	1	?	?	?	?	?	?	?	?	?	?	?	?	?
29 GREEN	0	1	1	?	?	?	?	?	?	?	?	?	?	?	?
30 CUMBERLAND	0	1	1	?	?	?	?	?	?	?	?	?	?	?	?
31 DUCK	0	1	1	?	?	?	?	0	1	0	0	0	0	0	0
32 LOWER TENNESSEE	?	?	?	1	0	?	0	0	?	0	0	?	?	?	?
33 UPPER TENNESSEE	?	1	1	?	?	1	?	0	?	0	0	?	?	?	?
34 MOBILE BASIN	?	0	0	1	0	0	0	?	0	?	?	0	0	0	0

[a] Matrix for 15 informative nodes of seven subtrees (fig. 23), assumption 2.

2. increased resolution within one grouping (areas 1–27) with addition of one novel node (areas 2, 4);

3. different grouping of the same areas (areas 32, 34).

PARSIMONY ANALYSIS OF NODES
AND SPECIES DISTRIBUTIONS
COMBINED FISH CLADOGRAMS (FIG. 25)

In his published matrix, Mayden included a binary character describing the distribution of each species (table 16, species 1–29). These
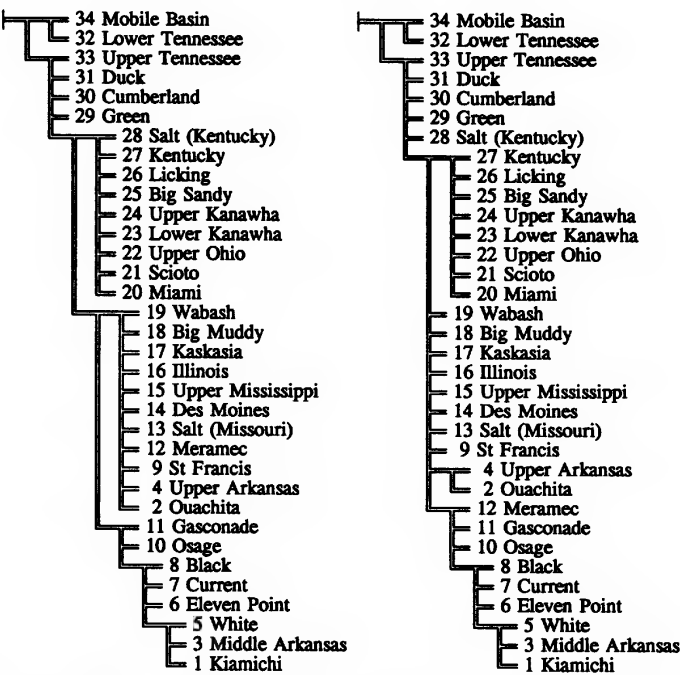
Left tree:
```
 34 Mobile Basin
 32 Lower Tennessee
 33 Upper Tennessee
 31 Duck
 30 Cumberland
 29 Green
 28 Salt (Kentucky)
 27 Kentucky
 26 Licking
 25 Big Sandy
 24 Upper Kanawha
 23 Lower Kanawha
 22 Upper Ohio
 21 Scioto
 20 Miami
 19 Wabash
 18 Big Muddy
 17 Kaskasia
 16 Illinois
 15 Upper Mississippi
 14 Des Moines
 13 Salt (Missouri)
 12 Meramec
  9 St Francis
  4 Upper Arkansas
  2 Ouachita
 11 Gasconade
 10 Osage
  8 Black
  7 Current
  6 Eleven Point
  5 White
  3 Middle Arkansas
  1 Kiamichi
```

Right tree:
```
 34 Mobile Basin
 32 Lower Tennessee
 33 Upper Tennessee
 31 Duck
 30 Cumberland
 29 Green
 28 Salt (Kentucky)
 27 Kentucky
 26 Licking
 25 Big Sandy
 24 Upper Kanawha
 23 Lower Kanawha
 22 Upper Ohio
 21 Scioto
 20 Miami
 19 Wabash
 18 Big Muddy
 17 Kaskasia
 16 Illinois
 15 Upper Mississippi
 14 Des Moines
 13 Salt (Missouri)
  9 St Francis
  4 Upper Arkansas
  2 Ouachita
 12 Meramec
 11 Gasconade
 10 Osage
  8 Black
  7 Current
  6 Eleven Point
  5 White
  3 Middle Arkansas
  1 Kiamichi
```

Fig. 24. Results of parsimony analysis of component matrix (table 13, assumption 2) for 15 nodes of subtrees (fig. 23). **Left** (nine nodes), one of four least resolved (table 14) of 5299+ (overflow) trees (length 15, ci 100, ri 100); of the four trees this one is least informative (table 15, 1340 independent three-item statements). **Right** (eight nodes), minimal tree (length 15) derived from least resolved (left) tree by collapse of nodes and more basal placement of some areas. The same minimal tree results from parsimony analysis of three-item matrices (see text).

binary characters, and multistate characters describing nodes of cladograms of the seven species groups (table 20, MS4–18) constitute his matrix. The two classes of characters dif-

TABLE 14

Number of Nodes of 5299 Trees Derived from Mayden's Data[a]

Number of		Number of	
Nodes	Trees	Nodes	Trees
28	64	18	342
27	184	17	324
26	222	16	247
25	184	15	264
24	147	14	124
23	129	13	51
22	764	12	37
21	858	11	32
20	765	10	19
19	538	9	4

[a] Resulting from parsimony analysis of component matrix for subtree nodes (table 13). See also table 15.

fer. Those for nodes of cladograms have ?-entries for missing data; those for species do not. Missing data reflect absence of particular areas from distributions of species related by a particular cladogram (the cladogram is seen as an individual subtree).

Binary characters describing distributions of species may be seen in two ways: (1) as presence/absence data independent of any tree; (2) as representations of terminal nodes of a tree, relating organisms of each species.

TABLE 15

Number of Independent Three-Item Statements[a]

Number of	
Statements	Trees
1508	1
1458	1
1442	1
1340	1

[a] Of the four least-resolved trees (table 14, nine nodes).

Left cladogram:

- 34 Mobile Basin
- 33 Upper Tennessee
- 32 Lower Tennessee
- 31 Duck
- 12 Meramec
- 11 Gasconade
- 10 Osage
- 8 Black
- 7 Current
- 6 Eleven Point
- 5 White
- 9 St Francis
- 4 Upper Arkansas
- 3 Middle Arkansas
- 2 Ouachita
- 1 Kiamichi
- 30 Cumberland
- 29 Green
- 28 Salt (Kentucky)
- 27 Kentucky
- 26 Licking
- 25 Big Sandy
- 24 Upper Kanawha
- 23 Lower Kanawha
- 22 Upper Ohio
- 21 Scioto
- 20 Miami
- 19 Wabash
- 18 Big Muddy
- 17 Kaskasia
- 16 Illinois
- 15 Upper Mississippi
- 14 Des Moines
- 13 Salt (Missouri)

Right cladogram:

- 34 Mobile Basin
- 33 Upper Tennessee
- 31 Duck
- 30 Cumberland
- 29 Green
- 19 Wabash
- 9 St Francis
- 12 Meramec
- 11 Gasconade
- 10 Osage
- 8 Black
- 7 Current
- 6 Eleven Point
- 5 White
- 4 Upper Arkansas
- 2 Ouachita
- 3 Middle Arkansas
- 1 Kiamichi
- 32 Lower Tennessee
- 27 Kentucky
- 22 Upper Ohio
- 21 Scioto
- 20 Miami
- 26 Licking
- 25 Big Sandy
- 23 Lower Kanawha
- 24 Upper Kanawha
- 28 Salt (Kentucky)
- 18 Big Muddy
- 17 Kaskasia
- 16 Illinois
- 15 Upper Mississippi
- 14 Des Moines
- 13 Salt (Missouri)

Fig. 25. Results of parsimony analysis of complete component matrices (combined matrices of tables 10 and 16; 28 nodes and 29 species distributions) for combined cladograms (figs. 18, 22). Left (18 nodes, no assumption 2), strict consensus of two trees (length 124, ci 45, ri 80). Right (23 nodes, assumption 2), one tree (length 122, ci 46, ri 79).

Seen as a terminal node, 1-entries of each binary character are related together relative to 0-entries. For the fish distributions, such a binary character without missing data includes 0-entries that result from comparison across paralogous nodes of the combined cladograms. Binary characters without missing data (table 16) may be deemed appropriate representations of terminal nodes of combined cladograms that include paralogous nodes (figs. 18, 22). Such binary characters are not appropriate for terminal nodes of individual subtrees (figs. 20, 23), because no subtree relates species occurring in all areas. As representations of terminal nodes, all such binary characters (table 16) include paralogous 0-entries.

Data for all nodes (table 10) and all species distributions (table 16) can be combined in a "complete" component matrix with maximum paralogy. Parsimony analysis of this matrix without assumption 2 yields two trees, with a strict consensus of 18 informative nodes (fig. 25, left). Parsimony analysis of the matrix with assumption 2 yields one tree of 23 nodes (fig. 25, right).

Comparison of figure 25 left and right shows the effects of the application of assumption 2 on results of parsimony analysis of a complete component matrix for the combined cladograms (nodes and species distributions):

1. overall increase in resolution (from 18 to 23 informative nodes);
2. increased resolution within one grouping (20–27), with addition of two novel nodes (areas 20–23, 25–27; areas 20–22, 27);
3. many different groupings of the same areas.

These effects contrast with those observed in comparison of results achieved only for nodes (fig. 19), where there are overall reduction in resolution, one novel node (23–36), and fewer different groupings of the same areas.

INDIVIDUAL FISH SUBTREES (FIG. 26)

Paralogy may be eliminated from the binary characters describing species distribu-

TABLE 16
Binary Matrix for Geographic Distribution of Fish Species[a]

Species:	0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2 2 2 2 2 2 2
	1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9
OUTGROUP	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
01 KIAMICHI	0 0 1 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
02 OUACHITA	0 0 1 0 0 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 1
03 MIDDLE ARKANSAS	0 0 1 0 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 1 1 0
04 UPPER ARKANSAS	0 0 1 0 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0
05 WHITE	0 0 1 0 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0
06 ELEVEN POINT	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 1 0 0 0 1 0 0 0
07 CURRENT	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0
08 BLACK	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 1 0 0 1 0 0 0
09 ST FRANCIS	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0
10 OSAGE	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0
11 GASCONADE	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0
12 MERAMEC	0 0 1 0 0 1 0 0 0 1 0 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0
13 SALT (MISSOURI)	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
14 DES MOINES	0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
15 UPPER MISSISSIPPI	0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
16 ILLINOIS	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
17 KASKASIA	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
18 BIG MUDDY	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
19 WABASH	0 0 0 0 0 0 0 0 0 1 0 0 0 1 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0
20 MIAMI	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
21 SCIOTO	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
22 UPPER OHIO	0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0
23 LOWER KANAWHA	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
24 UPPER KANAWHA	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
25 BIG SANDY	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
26 LICKING	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
27 KENTUCKY	0 0 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0
28 SALT (KENTUCKY)	0 0 0 0 0 0 0 0 0 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
29 GREEN	0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
30 CUMBERLAND	0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
31 DUCK	0 1 0 0 0 0 0 0 1 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
32 LOWER TENNESSEE	0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 0 0 1 0 0 0 0 0 1 0 0 0 0 0
33 UPPER TENNESSEE	0 1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 1 0 0 0 0 0 0 0 0 0 0 0
34 MOBILE BASIN	1 0 0 1 0 0 0 0 0 0 0 1 0 0 1 0 0 0 0 0 0 1 0 0 0 0 0 0 0

[a] For geographic distributions of 29 species (after Mayden, 1988: 335-337, table 1, see note below), no assumption 2. For assumption 2, certain 1-entries become 0-entries: area 3, species 28; area 4, species 8; area 32, species 2, 13, 24; area 33, species 5. Note: In his matrix Mayden treated species 13 and 14 as one species; they are here treated as separate species, save for the analysis of his corrected matrix (fig. 29, right).

tions by replacing certain 0-entries by ?-entries for missing data (table 17) for both series of individual subtrees considered above (figs. 20, 23). Revised characters for species distributions combine with characters for nodes to make two complete matrices: with and without assumption 2. Parsimony analysis of each complete subtree matrix yields one minimal tree (fig. 26, left and right). As rendered by subtree analysis, these minimal trees are offered as exact results for the paralogy-free

fraction of the complete data, including species distributions as if the distributions were terminal nodes of individual subtrees.

Comparison of figure 26 left and right shows effects of application of assumption 2 on results of parsimony analysis of a complete matrix for subtrees (nodes and species distributions seen as terminal nodes of individual subtrees):

1. overall increase in resolution (from 11 to 13 informative nodes);

Fig. 26. Minimal trees from parsimony analysis of complete component matrices for subtrees (clado-gram nodes and species distributions). **Left** (11 nodes), no assumption 2. **Right** (13 nodes), assumption 2. Complete matrix for left tree (CML) combines matrices for 15 nodes (table 11) and 19 species distributions (table 17, left, species 2–28). Complete matrix for right tree (CMR) combines matrices for 15 nodes (table 13) and 16 species distributions (table 17, right, species 2–26). Parsimony analysis of CML yields 5303+ (overflow) trees (length 39, ci 87, ri 95). Their strict consensus has three informative nodes (areas 1–33; areas 1–4; areas 20–27). Of 5303 trees, 10 are least resolved (17 nodes, range 17–27), of which two are least informative (2234 independent three-item statements, range 2234–2372). The two trees differ in placement of area 5 (either with areas 1–4 or with areas 6–9). The strict consensus of the two trees has length 40. With collapse of nodes and more basal placement of some areas, the consensus yields the left Minimal Tree (length 40). Parsimony analysis of CMR yields 5298+ (overflow) trees (length 38, ci 81, ri 92). Their strict consensus has one informative node (areas 2, 4). Of 5298 trees, 7 are least resolved (16 nodes, range 16–31), of which 1 tree is least informative (1920 independent 3-item statements, range 1920–2352). With collapse of nodes and more basal placement of some areas this tree yields the right Minimal Tree (length 38).

2. increased resolution within one grouping (areas 1–27) with additional of 1 novel node (areas 2, 4);
3. different groupings of the same areas (areas 32, 34).

These effects are similar to those observed in comparison of results achieved only for nodes (fig. 21, right; fig. 24, right).

DISCUSSION OF RESULTS OF PARSIMONY ANALYSES

AREAS WITH IDENTICAL CHARACTER STRINGS

There is variety in results of the above analyses. There is commonality as well, some of which stems from identical data for different areas. In all component matrices for nodes, certain areas have identical character strings, and parsimony analysis consequently places them together: areas 6–8; areas 10–11; areas 13, 16–18; areas 14–15; areas 20–22; areas 23–26; areas 29–30. The same applies to all component matrices for nodes and species distributions, with exception of areas 23–26, which are reduced to areas 23, 25–26.

Areas of any one of these seven groups seldom appear themselves as the only areas related by a node in trees resulting from parsimony analysis, but among 95 nodes of those trees, 10 (11%) relate areas only of one of the seven groups. For matrices of nodes, none of

TABLE 17
Component Matrices for Informative Distributions of Fish Species[a]

Species:	0 0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2 2 2	0 0 0 0 0 1 1 1 1 1 1 1 2 2 2 2
	2 3 5 6 8 9 0 1 3 4 7 8 9 1 2 4 5 6 8	2 3 6 8 9 0 1 3 4 7 8 9 1 2 5 6
OUTGROUP	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0
01 KIAMICHI	0 1 0 0 1 0 0 1 0 1 ? ? ? ? ? ? ? ? ?	0 1 0 1 0 0 1 0 1 ? ? ? ? ? ? ?
02 OUACHITA	0 1 ? ? ? 0 0 1 0 1 1 ? ? ? ? 0 ? ? 0	0 1 ? ? 0 0 1 0 1 1 ? ? ? ? 0 0
03 MIDDLE ARKANSAS	0 1 0 0 1 0 0 1 0 1 ? ? ? ? ? 0 0 0 1	0 1 0 1 0 0 1 0 1 ? ? ? ? ? 0 0
04 UPPER ARKANSAS	0 1 0 0 1 0 0 1 ? ? ? ? ? ? ? 0 ? ? 1	0 1 ? ? 0 0 1 ? ? ? ? ? ? ? 0 0
05 WHITE	0 1 0 0 0 0 1 0 0 1 1 0 0 0 1 0 0 1 ?	0 1 0 0 0 1 0 0 1 1 0 0 0 1 0 1
06 ELEVEN POINT	0 1 0 1 0 0 1 0 0 1 1 0 0 0 1 0 0 1 ?	0 1 1 0 0 1 0 0 1 1 0 0 0 1 0 1
07 CURRENT	0 1 0 1 0 0 1 0 0 1 1 0 0 0 1 0 0 1 ?	0 1 1 0 0 1 0 0 1 1 0 0 0 1 0 1
08 BLACK	0 1 0 1 0 0 1 0 0 1 1 0 0 0 1 0 0 1 ?	0 1 1 0 0 1 0 0 1 1 0 0 0 1 0 1
09 ST FRANCIS	0 1 0 1 0 0 1 0 0 1 1 ? ? ? ? ? ? ? ?	0 1 1 0 0 1 0 0 1 1 ? ? ? ? ? ?
10 OSAGE	0 1 0 1 0 0 1 0 0 1 ? 0 0 1 0 0 1 0 ?	0 1 1 0 0 1 0 0 1 ? 0 0 1 0 1 0
11 GASCONADE	0 1 0 1 0 0 1 0 0 1 ? 0 0 1 0 0 1 0 ?	0 1 1 0 0 1 0 0 1 ? 0 0 1 0 1 0
12 MERAMEC	0 1 0 1 0 0 1 0 0 1 ? 0 0 1 0 ? ? ? ?	0 1 1 0 0 1 0 0 1 ? 0 0 1 0 ? ?
13 SALT (MISSOURI)	? ? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ? ? ?	? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ?
14 DES MOINES	0 1 ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ? ? ?	0 1 ? ? 0 1 0 ? ? ? ? ? ? ? ? ?
15 UPPER MISSISSIPPI	0 1 ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ? ? ?	0 1 ? ? 0 1 0 ? ? ? ? ? ? ? ? ?
16 ILLINOIS	? ? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ? ? ?	? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ?
17 KASKASIA	? ? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ? ? ?	? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ?
18 BIG MUDDY	? ? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ? ? ?	? ? ? ? 0 1 0 ? ? ? ? ? ? ? ? ?
19 WABASH	? ? ? ? ? 0 1 0 0 1 1 ? ? ? ? ? ? ? ?	? ? ? ? 0 1 0 0 1 1 ? ? ? ? ? ?
20 MIAMI	? ? ? ? ? 0 1 0 ? ? 0 1 0 0 ? ? ? ? ?	? ? ? ? 0 1 0 ? ? 0 1 0 0 ? ? ?
21 SCIOTO	? ? ? ? ? 0 1 0 ? ? 0 1 0 0 ? ? ? ? ?	? ? ? ? 0 1 0 ? ? 0 1 0 0 ? ? ?
22 UPPER OHIO	? ? ? ? ? 0 1 0 ? ? 0 1 0 0 ? ? ? ? ?	? ? ? ? 0 1 0 ? ? 0 1 0 0 ? ? ?
23 LOWER KANAWHA	? ? ? ? ? ? ? ? ? ? 0 1 0 0 ? ? ? ? ?	? ? ? ? ? ? ? ? ? 0 1 0 0 ? ? ?
24 UPPER KANAWHA	? ? ? ? ? ? ? ? ? ? 0 0 0 0 ? ? ? ? ?	? ? ? ? ? ? ? ? ? 0 0 0 0 ? ? ?
25 BIG SANDY	? ? ? ? ? ? ? ? ? ? 0 1 0 0 ? ? ? ? ?	? ? ? ? ? ? ? ? ? 0 1 0 0 ? ? ?
26 LICKING	? ? ? ? ? ? ? ? ? ? 0 1 0 0 ? ? ? ? ?	? ? ? ? ? ? ? ? ? 0 1 0 0 ? ? ?
27 KENTUCKY	? ? ? ? ? 0 1 0 1 0 ? 0 1 0 0 ? ? ? ?	? ? ? ? 0 1 0 1 0 ? 0 1 0 0 ? ?
28 SALT (KENTUCKY)	? ? ? ? ? ? ? 1 0 ? ? ? ? ? ? ? ? ? ?	? ? ? ? ? ? 1 0 ? ? ? ? ? ? ? ?
29 GREEN	1 0 ? ? ? 1 0 0 1 0 ? ? ? ? ? ? ? ? ?	1 0 ? ? 1 0 0 1 0 ? ? ? ? ? ? ?
30 CUMBERLAND	1 0 ? ? ? 1 0 0 1 0 ? ? ? ? ? ? ? ? ?	1 0 ? ? 1 0 0 1 0 ? ? ? ? ? ? ?
31 DUCK	1 0 ? ? ? 1 0 0 1 0 ? 1 0 0 0 1 0 0 0	1 0 ? ? 1 0 0 1 0 ? 1 0 0 0 0 0
32 LOWER TENNESSEE	1 0 1 0 0 ? ? ? 1 0 ? 1 0 0 0 1 0 0 0	? ? 0 0 ? ? ? ? ? ? 1 0 0 0 ? ?
33 UPPER TENNESSEE	1 0 1 0 0 ? ? ? 1 0 0 1 0 0 0 ? ? ? ?	1 0 ? ? ? ? ? 1 0 0 1 0 0 0 ? ?
34 MOBILE BASIN	0 0 0 0 0 ? ? ? 0 0 0 ? ? ? ? 0 0 0 0	0 0 0 0 ? ? ? 0 0 0 ? ? ? ? 0 0

[a] For informative geographic distributions of 29 species considered as terminal nodes of subtrees: left (species 2–28), no assumption 2 (fig. 20); right (species 2–26), assumption 2 (fig. 23).

the seven groups has any unique character; for matrices including species distributions, only one (10–11) has. Otherwise, areas of one or another of the seven groups appear related by a node because of optimization of homoplastic characters. As such, these nodes are possibly artifactual. If the resulting trees are divided into two classes according to the nature of their matrix (combined cladograms versus individual subtrees), then 9 of the 10 nodes occur in the combined class (16% of 57 nodes), and only one appears in the individual class (3% of 38 nodes). These findings suggest that paralogy, which is maximal in matrices for combined cladograms and absent from matrices for individual subtrees, results in artifactual nodes.

EFFECTS OF PARALOGY

Effects of paralogy may be assessed by comparison of the consistency index of trees derived by parsimony analysis of the two classes of matrices (table 18). The consistency index of trees derived by parsimony analysis of matrices for individual subtrees is higher by a factor of two. Number of nodes resolved is lower by a third. These findings suggest that

TABLE 18
Consistency Index (ci) and Number of Nodes of Trees[a]

Matrix class	Nodes only		Nodes and species distributions		Average ci	Total nodes
	no A2	A2	no A2	A2		
Combined						
ci	38	37	45	46	42	
Nodes	10	6	18	23		57
Individual						
ci	93	100	87	81	90	
Nodes	6	8	11	13		38

[a] Derived from two classes of matrices (combined and individual). Results for combined matrices are shown in figures 19, left and right, and 25, left and right. Results for individual matrices are shown in figures 21, right; 24, right; and 26, left and right. Abbreviations: no A2, no assumption 2; A2, assumption 2.

paralogy contributes both artifactual inconsistency to matrices for combined cladograms and artifactual nodes to results of parsimony analysis of these matrices.

EFFECTS OF ASSUMPTION 2

Effects of assumption 2 are not apparent by such comparison (table 18). In matrices for combined cladograms, assumption 2 causes some 1-entries to become 0-entries (data change, and paralogy is apt to increase as a result). For results of parsimony analysis of such matrices, assumption 2 hardly affects consistency and causes both decrease (nodes only) and increase (nodes and species distributions) in nodes resolved. In matrices for individual subtrees, assumption 2 causes some 1-and 0-entries to become ?-entries (conflicting data disappear). For results of parsimony analysis of such matrices, assumption 2 causes both increase (nodes only) and decrease (nodes and species distributions) in consistency, and in both cases causes increase in number of nodes resolved. The decrease in consistency (from 87 to 81, nodes and species distributions) may seem anomalous. It may be remembered, however, that assumption 2 applies not to conflicts among species distributions (seen as terminal nodes)
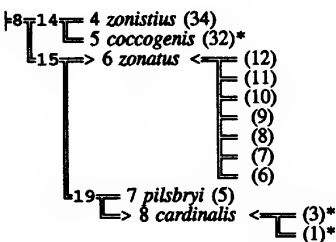
Fig. 27. Cladogram (node 8) in which widespread species (*zonatus* and *cardinalis*) are seen as terminal nodes (cf. fig. 22).

but only to conflicts arising within and between the 27 nodes of the combined cladograms, and then only to conflicts involving endemic and widespread species (figs. 18, 22).

The complete matrix for assumption 2 (table 13 for nodes, table 17: 2–26 for species distributions) includes six characters that reduce consistency of the resulting tree (table 19). If all six characters are rendered inactive, then the consistency (ci) of the resulting tree is 100. Of the six conflicting species distributions, assumption 2 might be applied only to one of them (species 18), with reduction of its distribution to areas 31 and 33 (the relationship of area 32 is determined by the possible endemic, *L. coccogenis*; see above). The resulting consistency (83) would still be a reduction (from 87 for the result without assumption 2). Distributions of the remaining five species might, of course, be reduced to yield 100% consistency, but not by application of assumption 2 (there are no endemic species relevant to such reduction).

Treating species distributions as terminal nodes tends to render absurd notions of endemic and widespread species. Consider, for example, the cladogram of node 8 (fig. 27), with widespread species *zonatus* and *cardinalis* treated as terminal nodes. The resulting terminals are merely areas with organisms that lack any distinguishing biological features. All terminals become single areas as if each were characterized by an endemic taxon. If the node *zonatus* were reduced in a manner analogous with application of assumption 2, so as to relate organisms only of areas 6–8, or of areas 10–12, then the reduced *zonatus* would be consistent with results of parsimony analysis of the entire matrix (fig. 26, right). The object of consistency aside, there is no

rationale to govern such reduction. Assumption 2 is inapplicable, therefore, to five of the six species distributions.

The import of assumption 2 lies in another dimension: identification of widespread species of which the distribution conflicts with area relationship as determined by endemics. In the case of the fishes, conflicts identified concern areas 1–4 and areas 31–34. These two groups of areas involve major changes in drainage pattern, particularly area 4 (Upper Arkansas) and area 32 (Lower Tennessee). At an earlier time, the "Upper Arkansas" was part of the Plains Stream, which entered the Gulf of Mexico independently of the Mississippi system; the "Lower Tennessee" was part of the Old Tennessee River, with a possible connection to the Mobile Basin (Mayden, 1988: fig. 4). Application of assumption 2 associates the Upper Arkansas and Ouachita (areas 2 and 4), and the Lower Tennessee and Mobile Basin (areas 32 and 34); identifies possible endemics of these older drainages (*E. cragini* to the Upper Arkansas, *L. coccogenis* to the Lower Tennessee); and suggests that other species, occurring there today, are not native to the older drainages (*L. cardinalis* in the Upper Arkansas; *N. leuciodus*, *F. catenatus* east, and *E. boschungi* in the Lower Tennessee). These implications are testable, perhaps, only indirectly and with difficulty. Area relationships (areas 2 and 4; areas 32 and 34), however, are more directly testable through analysis of other taxa.

The significance of the six species, of which the distribution reduces consistency, lies in a yet different dimension. It is possible, of course, that each such species at one time was endemic to one area, and subsequently became widespread (all things are possible). The distribution of each species, when fitted to the tree resulting from parsimony analysis of the complete matrix (fig. 26, right), suggests another possibility—that each species is a complex of forms with diverse relationships. This implication is directly testable by further study of geographic variation of the widespread species with conflicting distributions.

One may note (table 19) that identification of these six species with conflicting distributions and assessment of their possible significance do not require the complete matrix

TABLE 19
Inconsistent Species Distribution of Freshwater Fishes[a]

Matrix character (1st = 0)	Species	Distribution	ci of result	
			fig. 26	fig. 24
17	6	6–12	83	94
20	10	5–22, 27	85	88
21	11	1–4	83	94
22	13	27–31, 33	83	94
25	18	31–33	83	94
30	26	5–8	83	94

[a] Inconsistent with results of parsimony analysis of matrices with (fig. 26, from table 13 + table 17: 2–26) and without (fig. 24, from table 13) binary characters for distribution of 29 species, assumption 2. For figure 26, the ci is that of resulting tree when given matrix character is rendered inactive (if active, ci = 81). For figure 24, the ci is that of the resulting tree when character is rendered active (if inactive ci = 100). For figure 26, the six characters are the only characters with length greater than one (characters 17, 21, 22, 25, and 30, length 2; 20, length 3). For figure 24, other characters (all inactive) have length greater than one (characters 15, 17–19, 23, 25–27, 29, and 30, length 2; 21 and 22, length 3; and 20, length 4); the characters with length greater than one and not included in table (15, 18, 19, 23, 26, 27, and 29) are each compatible with the tree such that if the character is rendered active, consistency is 100% (the resulting tree is different from that of fig. 24).

(nodes and species distributions; table 13 + table 17: 2–26) nor results of parsimony analysis of this matrix (fig. 26). Identification of species with conflicting distributions may be achieved with the matrix for nodes only (table 13) and their significance assessed by means of the tree (fig. 24) resulting from parsimony analysis of that matrix.

Earlier, we were skeptical of the value of including in a matrix characters for distributions of species, which we saw as mere presence/absence ("phenetic") data independent of any tree (Nelson and Ladiges, 1991a, 1991b). Viewed as terminal nodes of trees, however, species distributions are cast in a different light. Nevertheless, we remain skeptical, and unconvinced, of their value. In the present case, they add little or nothing to results obtainable by parsimony analysis of nodes only (cf. figs. 21, 26, left; 24, 26, right).

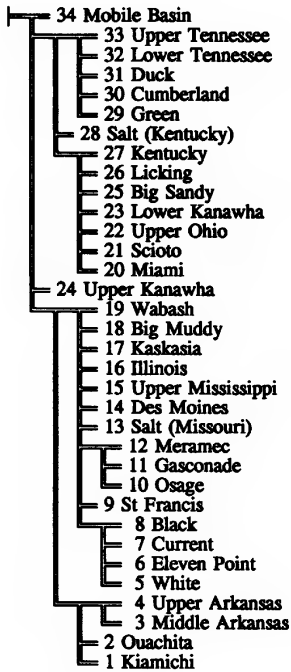If species distributions (seen as terminal nodes of individual subtrees) are by them-

```
┌── 34 Mobile Basin
│  ┌─ 33 Upper Tennessee
│  ├─ 32 Lower Tennessee
│  ├─ 31 Duck
│  ├─ 30 Cumberland
│  └─ 29 Green
├─ 28 Salt (Kentucky)
│  ┌─ 27 Kentucky
│  ├─ 26 Licking
│  ├─ 25 Big Sandy
│  ├─ 23 Lower Kanawha
│  ├─ 22 Upper Ohio
│  ├─ 21 Scioto
│  └─ 20 Miami
├─ 24 Upper Kanawha
│  ┌─ 19 Wabash
│  ├─ 18 Big Muddy
│  ├─ 17 Kaskasia
│  ├─ 16 Illinois
│  ├─ 15 Upper Mississippi
│  ├─ 14 Des Moines
│  ├─ 13 Salt (Missouri)
│  │  ┌─ 12 Meramec
│  │  ├─ 11 Gasconade
│  │  └─ 10 Osage
│  ├─ 9 St Francis
│  ├─ 8 Black
│  ├─ 7 Current
│  ├─ 6 Eleven Point
│  └─ 5 White
│  ┌─ 4 Upper Arkansas
│  └─ 3 Middle Arkansas
├─ 2 Ouachita
└─ 1 Kiamichi
```

←

Fig. 28. Minimal tree (nine nodes) from parsimony analysis of component matrix for 19 species distributions (table 17, left, species 2–28, no assumption 2). Parsimony analysis yields 5302+ (overflow) trees (length 21, ci 90, ri 96). Their strict consensus has three informative nodes (areas 1–19, 24; areas 3–4; areas 20–23, 25–33). Of 5302 trees, 2 are least resolved (13 nodes, range 13–28), of which 1 is least informative (1908 independent three-item statements, range 1908–1968). Collapse of nodes and more basal placement of some areas yield this minimal tree (length 21). With assumption 2 (table 17, right, species 2–26), the minimal tree (eight nodes, length 18, ci 88, ri 96) is the same but lacks a node relating areas 3–4.

Left tree:
```
┌── 34 Mobile Basin
│  ┌─ 33 Upper Tennessee
│  ├─ 32 Lower Tennessee
│  ├─ 31 Duck
│  ├─ 30 Cumberland
│  └─ 29 Green
├─ 28 Salt (Kentucky)
│  ┌─ 27 Kentucky
│  │  ┌─ 22 Upper Ohio
│  │  ├─ 21 Scioto
│  │  └─ 20 Miami
│  ├─ 26 Licking
│  ├─ 25 Big Sandy
│  ├─ 23 Lower Kanawha
│  └─ 24 Upper Kanawha
├─ 18 Big Muddy
├─ 17 Kaskasia
├─ 16 Illinois
├─ 13 Salt (Missouri)
│  ┌─ 15 Upper Mississippi
│  ├─ 14 Des Moines
│  │  ┌─ 12 Meramec
│  │  ├─ 11 Gasconade
│  │  └─ 10 Osage
│  └─ 19 Wabash
├─ 9 St Francis
├─ 8 Black
├─ 7 Current
├─ 6 Eleven Point
├─ 5 White
│  ┌─ 4 Upper Arkansas
│  └─ 3 Middle Arkansas
├─ 1 Kiamichi
└─ 2 Ouachita
```

Right tree:
```
┌── 34 Mobile Basin
│  ┌─ 33 Upper Tennessee
│  ├─ 32 Lower Tennessee
│  ├─ 31 Duck
│  ├─ 30 Cumberland
│  └─ 29 Green
├─ 28 Salt (Kentucky)
│  ┌─ 27 Kentucky
│  │  ┌─ 22 Upper Ohio
│  │  ├─ 21 Scioto
│  │  └─ 20 Miami
│  ├─ 26 Licking
│  ├─ 25 Big Sandy
│  ├─ 23 Lower Kanawha
│  └─ 24 Upper Kanawha
├─ 18 Big Muddy
├─ 17 Kaskasia
├─ 16 Illinois
├─ 13 Salt (Missouri)
│  ┌─ 15 Upper Mississippi
│  ├─ 14 Des Moines
│  │  ┌─ 12 Meramec
│  │  ├─ 11 Gasconade
│  │  └─ 10 Osage
│  └─ 19 Wabash
├─ 9 St Francis
├─ 8 Black
├─ 7 Current
├─ 6 Eleven Point
├─ 5 White
│  ┌─ 4 Upper Arkansas
│  └─ 3 Middle Arkansas
├─ 2 Ouachita
└─ 1 Kiamichi
```

Fig. 29. Strict consensus trees. **Left** (16 nodes), tree obtained by Mayden (1988: 344), confirmed with slight correction of matrix by Nelson and Ladiges (1991a: 44, 57). **Right** (15 nodes), tree obtained by parsimony analysis of Mayden's matrix, with correction of characters describing cladograms (table 20, right, TS4–27). For the full matrix there are 5304+ (overflow) trees (length 143, ci 81, ri 89). For the consolidated matrix (areas with duplicate character strings deleted), there are 165 trees (length 143, ci 81, ri 85). Two-state characters for nodes weighted ×4 (multistate characters for nodes weighted ×2 in Mayden's original analysis). Lower differential weights yield less resolution.
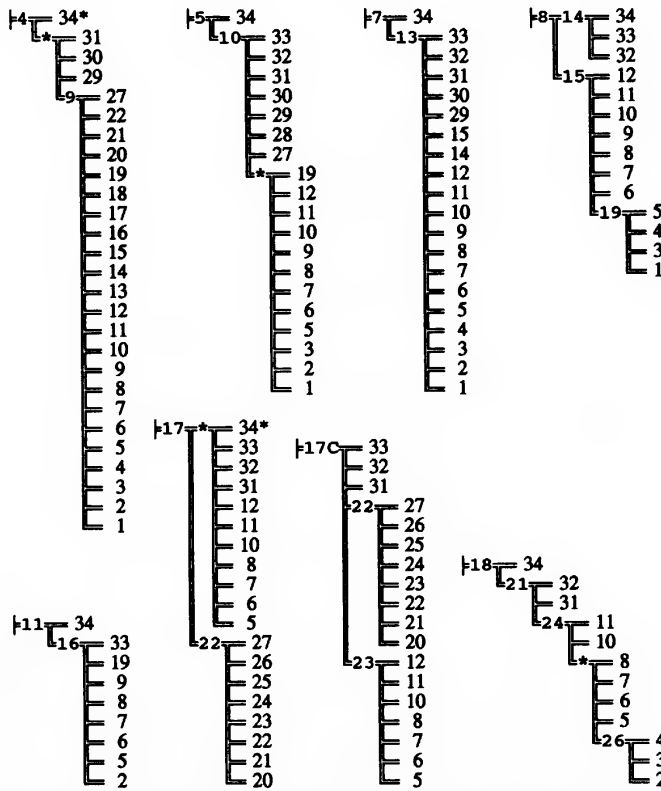
Fig. 30. Results of parsimony analysis of each of Mayden's multistate characters (table 20, left, MS4–18) describing seven cladograms of fishes. Cladograms (nodes) 4, 5, 17, 18 are erroneous (*). Cladogram (node) 17C partially corrected by Nelson and Ladiges (1991a; absent is node 20, uniting nodes 22 and 23). Errors: node 4, area 34 erroneously included; node 5, erroneous node (areas 1–19); node 17, erroneous node (areas 5–34), area 34 erroneously included, no node 20 (uniting nodes 22 and 23); node 18, erroneous node (areas 2–8), no node 25 (areas 3–11), no node 27 (areas 3–8).

selves combined in a component matrix (table 17: 2–28 or 2–26) then parsimony analysis of this matrix yields a minimal tree (fig. 28, nine nodes) with branching at variance with that of trees obtained from parsimony analysis of matrices only for subtree nodes (figs. 21, right; 24, right). Conflict arises not from paralogy (matrices for species distributions and nodes are paralogy free). It seems merely futile, and false, to expect species distributions themselves to reflect the historical pattern of branching.

MAYDEN'S ORIGINAL MATRIX (FIG. 29)

With parsimony analysis of a complete matrix (including a binary character representing the distribution of each of the 29 species [see below] and multistate characters for the nodes of the seven cladograms treated as individual subtrees), Mayden obtained 33 trees, of which the strict consensus (fig. 29, left) has 16 informative nodes (confirmed with slight correction of the matrix by Nelson and Ladiges, 1991a: 57).

As published by Mayden (1988: 336–377, table 1), the multistate characters (table 20, MS4–18; fig. 30) for four of the seven cladograms of these analyses are inaccurate. Nevertheless, parsimony analysis of the corrected matrix, with cladogram nodes represented by component data (table 20, TS4–27; fig. 31), yields more numerous trees, with a strict consensus (fig. 29, right, 15 informative nodes) only slightly different from Mayden's result.

Aside from effects of assumption 2, results obtained by Mayden (fig. 29, original and corrected) are not very different from those achieved here for complete matrices (fig. 26). The reasons are easy to understand. His ma-
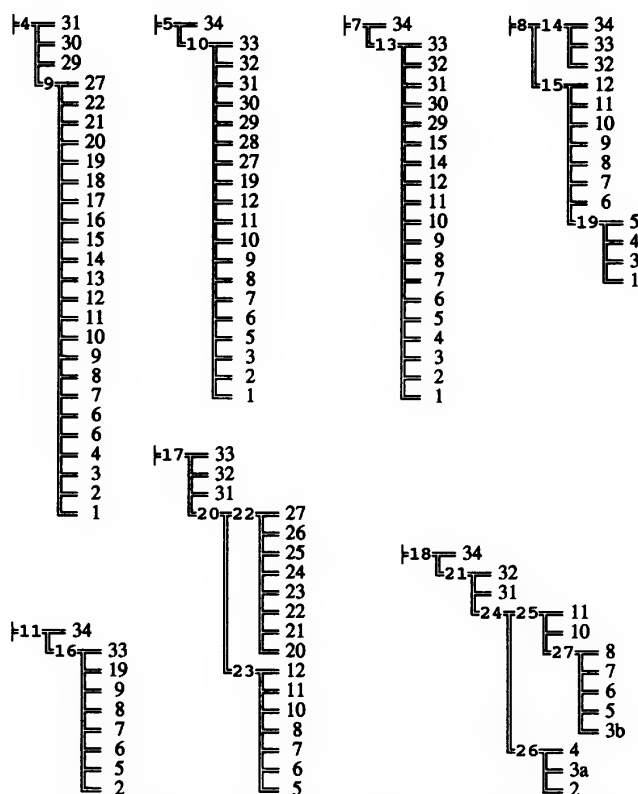
Fig. 31. Results of parsimony analysis of each series of two-state characters (table 20, right, TS4–27), correctly describing seven cladograms of fishes. For each series the consistency index (ci) is 100, except for that of cladogram (node) 18, in which area 3 appears twice. If the two occurrences of area 3 are distinguished (e.g., as 3a and 3b), then the result is as shown, with ci 100. If the two occurrences are not distinguished, then the result is as shown for area 3b, with ci 85.

trix includes characters for nodes as if nodes were of subtrees (characters are paralogy free). His matrix includes characters for species distributions as if distributions were terminal nodes of combined cladograms (characters are maximally paralogous). Computational problems are apparently overcome in this case (see below) by differential weighting of characters for nodes— $\times 2$ for multistates, equivalent to $\times 4$ for two-state equivalents (fig. 29).

ASSUMPTION ZERO

To include in a matrix characters for both cladogram nodes and species distributions is a practice begun by Wiley (1987: 297–299) in "an example of how a parsimony analysis might work, using the same hypothetical taxa as Humphries and Parenti (1986)," [not] "a formal method but only a sketch of the gen-

eral outlines of a formal method." Zandee and Roos (1987) formalized the practice under the name of "assumption zero." Wiley suggested that species distributions be represented as characters without missing data, and appropriate nodes of cladograms be represented as characters with missing data—a suggestion imperfectly implemented in his treatment of hypothetical examples (Wiley, 1987: table 3, a binary matrix wherein most, but not all, relevant cladogram nodes are represented by characters with missing data— column BX is the exception, corrected in his figure 10). "The binary coding for groups with incomplete distributional patterns is . . . complicated. There are three possibilities. . . . I coded all hypothetical ancestors under this condition as missing data" (Wiley, 1987: 301–302).

Wiley (1988a) later renamed the practice

TABLE 20
Multistate (MS) and Two-State (TS) Characters for Cladogram Nodes[a]

Nodes of Cladograms:	MS 0 0 0 0 1 1 1 1 1 / 4 5 7 8 1 7 7 8 8	TS 0 0 / 4 9	TS 0 1 / 5 0	TS 0 1 / 7 3	TS 0 1 1 1 / 8 4 5 9	TS 1 1 / 1 6	TS 1 2 2 2 / 7 0 2 3	TS 1 2 2 2 2 2 / 8 1 4 5 6 7
OUTGROUP	0 0 0 1 0 0 1 0 0	0 0	0 0	0 0	0 0 0 0	0 0	0 0 0 0	0 0 0 0 0 0
01 KIAMICHI	2 3 2 3 ? ? ? ? ?	1 1	1 1	1 1	1 0 1 1	? ?	? ? ? ?	? ? ? ? ? ?
02 OUACHITA	2 3 2 7 2 ? ? ? 4 ?	1 1	1 1	1 1	? ? ? ?	1 1	? ? ? ?	1 1 1 0 1 0
03 MIDDLE ARKANSAS	2 3 2 3 7 ? ? ? 4 3	1 1	1 1	1 1	1 0 1 1	? ?	? ? ? ?	1 1 1 1 1 1
04 UPPER ARKANSAS	2 ? 2 3 ? ? ? ? 4 ?	1 1	? ?	1 1	1 0 1 1	? ?	? ? ? ?	1 1 1 0 1 0
05 WHITE	2 3 2 3 2 1 0 3 2	1 1	1 1	1 1	1 0 1 1	1 1	1 1 0 1	1 1 1 1 0 1
06 ELEVEN POINT	2 3 2 2 2 1 0 3 2	1 1	1 1	1 1	1 0 1 0	1 1	1 1 0 1	1 1 1 1 0 1
07 CURRENT	2 3 2 2 2 1 0 3 2	1 1	1 1	1 1	1 0 1 0	1 1	1 1 0 1	1 1 1 1 0 1
08 BLACK	2 3 2 2 2 1 0 3 2	1 1	1 1	1 1	1 0 1 0	1 1	1 1 0 1	1 1 1 1 0 1
09 ST FRANCIS	2 3 2 2 2 ? ? ? ?	1 1	1 1	1 1	1 0 1 0	1 1	? ? ? ?	? ? ? ? ? ?
10 OSAGE	2 3 2 2 ? 1 0 3 1	1 1	1 1	1 1	1 0 1 0	? ?	1 1 0 1	1 1 1 1 0 0
11 GASCONADE	2 3 2 2 ? 1 0 3 1	1 1	1 1	1 1	1 0 1 0	? ?	1 1 0 1	1 1 1 1 0 0
12 MERAMEC	2 3 2 2 ? 1 0 ? ?	1 1	1 1	1 1	1 0 1 0	? ?	1 1 0 1	? ? ? ? ? ?
13 SALT (MISSOURI)	2 ? ? ? ? ? ? ? ?	1 1	? ?	? ?	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
14 DES MOINES	2 ? 2 ? ? ? ? ? ?	1 1	? ?	1 1	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
15 UPPER MISSISSIPPI	2 ? 2 ? ? ? ? ? ?	1 1	? ?	1 1	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
16 ILLINOIS	2 ? ? ? ? ? ? ? ?	1 1	? ?	? ?	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
17 KASKASIA	2 ? ? ? ? ? ? ? ?	1 1	? ?	? ?	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
18 BIG MUDDY	2 ? ? ? ? ? ? ? ?	1 1	? ?	? ?	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
19 WABASH	2 3 ? ? ? 2 ? ? ?	1 1	1 1	? ?	? ? ? ?	1 1	? ? ? ?	? ? ? ? ? ?
20 MIAMI	2 ? ? ? ? 1 2 ? ?	1 1	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
21 SCIOTO	2 ? ? ? ? 1 2 ? ?	1 1	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
22 UPPER OHIO	2 ? ? ? ? 1 2 ? ?	1 1	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
23 LOWER KANAWHA	? ? ? ? ? 1 2 ? ?	? ?	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
24 UPPER KANAWHA	? ? ? ? ? 1 3 ? ?	? ?	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
25 BIG SANDY	? ? ? ? ? 1 2 ? ?	? ?	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
26 LICKING	? ? ? ? ? 1 2 ? ?	? ?	? ?	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
27 KENTUCKY	2 2 ? ? ? 1 2 ? ?	1 1	1 1	? ?	? ? ? ?	? ?	1 1 1 0	? ? ? ? ? ?
28 SALT (KENTUCKY)	? 2 ? ? ? ? ? ? ?	? ?	1 1	? ?	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
29 GREEN	1 2 2 ? ? ? ? ? ?	1 0	1 1	1 1	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
30 CUMBERLAND	1 2 2 ? ? ? ? ? ?	1 0	1 1	1 1	? ? ? ?	? ?	? ? ? ?	? ? ? ? ? ?
31 DUCK	1 2 2 ? ? 1 0 2 0	1 0	1 1	1 1	? ? ? ?	? ?	1 0 0 0	1 1 0 0 0 0
32 LOWER TENNESSEE	? 2 2 0 ? 1 0 2 0	? ?	1 1	1 1	1 1 0 0	? ?	1 0 0 0	1 1 0 0 0 0
33 UPPER TENNESSEE	? 2 2 0 2 1 0 ? ?	? ?	1 1	1 1	1 1 0 0	1 1	1 0 0 0	? ? ? ? ? ?
34 MOBILE BASIN	0 1 1 0 1 ? 0 1 0	? ?	1 0	1 0	1 1 0 0	1 0	? ? ? ?	1 0 0 0 0 0

[a] Nodes 4–18 (left), MS characters for cladograms of seven species groups (after Mayden, 1988: 336–337, table 1). Nodes 4–27 (right), TS characters equivalent to corrected MS characters. MS4 is correctly represented by TS4 + 9; MS5, by TS5 + 10; MS7, by TS7 + 13; MS8, by TS8 + 14 + 15 + 19; MS11, by TS11 + 16; MS17, by TS17 + 20 + 22 + 23; MS18, by TS18 + 21 + 24 + 25 + 26 + 27. The correction by Nelson and Ladiges (1991: 57) replaces 0-entries by 1-entries for areas 31–33 (MS17, character 2); and one 0-entry by a ?-entry for area 34 (MS17, character 2).

as "Brooks Parsimony Analysis" (BPA) and maintained the different style of characters for species distributions and cladogram nodes (Wiley, 1988a: tables 3, 4; 1988b: table 3). The practice was continued by Wiley et al. (1991: table 7.4). Characters with missing data were extended to species distributions by Page (1990: 124, table 2) in order to make "explicit the relationship between the cladogram and the matrix," and without comment by Brooks (1990: tables 12, 14-15) and Funk and Brooks (1990: tables 9, 10). Nelson and Ladiges (1991b: 473) noted that "assumption zero treats a widespread taxon as if it were a node relating the areas in which the taxon occurs." Humphries (1992: fig. 9.15) depicted wide-
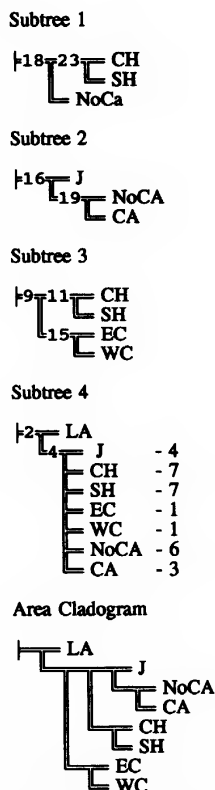
Subtree 1

```
 18  23   CH
          SH
          NoCa
```

Subtree 2

```
 16   J
     19   NoCA
          CA
```

Subtree 3

```
 9  11   CH
         SH
      15   EC
          WC
```

Subtree 4

```
 2   LA
    4   J      - 4
        CH     - 7
        SH     - 7
        EC     - 1
        WC     - 1
        NoCA   - 6
        CA     - 3
```

Area Cladogram

```
    LA
        J
            NoCA
            CA
            CH
            SH
        EC
        WC
```

Fig. 32. Subtrees and area cladogram for beetles of the genus *Platynus* (modified from Page and Lydeard, 1994: figs. 9, 10). Symbols: CA, Central America; CH, central and northern Hispaniola; EC, eastern Cuba; J, Jamaica; LA, Lesser Antilles; NoCA, northern Central America; SH, southern Hispaniola; WC, western Cuba. In Subtree 4 are numbers showing redundancy of areas distal to paralogous node 4 (see fig. 33).

spread species as terminal cladogram nodes and represented widespread species in a matrix by characters with missing data.

MISSING DATA

A matrix with missing data (e.g., table 13) invites existing parsimony programs to find trees that have shortest length but are overresolved for data of the matrix (see Methods). Of trees found for the matrix of table 13, for example, there are 64 trees with 28 informative nodes—for a matrix of 15 two-state characters and 100% consistency! In this case, programs did not find the minimal tree (eight nodes) before filling computer memory with overresolved trees.

Consider again Mayden's original matrix, which for species distributions includes characters without missing data. During parsimony analysis of this matrix, accumulation of overresolved trees in computer memory is inhibited. Optimization of characters for species distributions permits an informative result without search for a least resolved tree. For example, the first of 165 trees found for the consolidated matrix has 20 informative nodes. Of 28 species, 11 are endemic, leaving 17 widespread species. Characters for these 17 species distributions optimize (13 of the 17 as homoplasies) such that homoplasies define 14 of the 20 nodes of the tree. With differential weighting in favor of cladogram nodes, characters for species distributions inhibit accumulation of overresolved trees, with consequent lack of resolution in their strict consensus. This problem was alluded to by Wiley (1988b: 530):

> R. L. Mayden's . . . analysis of faunas of areas of endemism in the Central Highlands of North America includes a number of "relic" areas that contain only a small number of the total groups in the analysis. To counter false placement, he was forced to take these areas out of the initial analysis and treat them as one might treat a poorly preserved fossil, that is, adding them in a post hoc manner according to the evidence actually associated with them.

Wiley's "relic" areas refer, among others, to areas 13, 16–18, 23–26, and 28, each of which is part of the distribution only of 1 widespread species of the 29 fishes of Mayden's study and which therefore occurs only in one of the seven species groups. Areas 13 and 16–18, for example, are part of the distribution of *Nocomis biguttatus* and no other species (fig. 18, species 10); areas 23–26 are part of the distribution of *Etheostoma variatum* (fig. 18, species 19); area 28 is part of the distribution of *Fundulus catenatus* (fig. 18, species 13). As such these areas appear only as missing data in the component characters for the nodes of six of the seven species groups. The tactic described by Wiley would have as its effect the elimination of some missing data from the matrix, but that tactic was not used by Mayden in his published analysis. Deleterious effects of missing data were inhibited by binary characters for species distributions (see above) and, before parsimony analysis, by eliminating from the matrix areas with
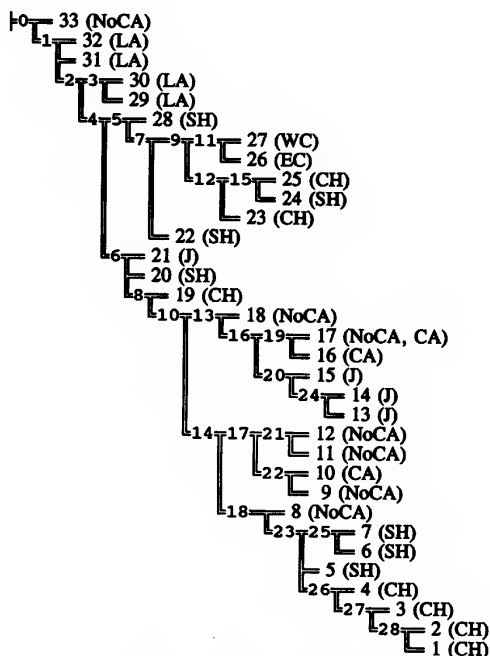
Fig. 33. Relationships of 33 species of beetles of the genus *Platynus* (modified from Page and Lydeard, 1994: fig. 8; after Liebherr, 1988: fig. 6). Symbols as in figure 32.
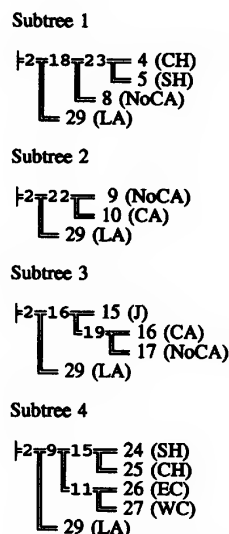

Fig. 34. Four subtrees derived from Liebherr's (1988) cladogram of beetles (fig. 33). Symbols as in figure 32.

duplicate character strings: "The number of equally parsimonious trees was reduced somewhat in this analysis by consolidating drainages with the same character strings" (Mayden, 1988: 335). Nelson and Ladiges (1991a: fig. 3) found that the full matrix yields about 1500 trees, and the matrix for consolidated drainages yields 33 trees—a reduction of about 98%.

GENERAL DISCUSSION

SUBTREES AND SUBTREES

In "Designing a Biogeographic Study," Page and Lydeard (1994) considered an empirical example, one of Liebherr's (1988) taxon-area cladograms for beetles of the genus *Platynus*. They noted:

if we decompose the cladogram into subtrees that minimize redundancy (this can be likened to identifying the sets of biogeographically orthologous taxa; see Nelson and Ladiges, 1991b: 481; and Page, 1993), we see that all the subtrees are mutually consistent . . . [fig. 32, Subtrees 1–4]; that is we could combine them all to create one or more area cladograms that

all subtrees could agree with. Figure . . . [32, Area Cladogram] shows the area cladogram for *Platynus* that has the fewest items of error (Nelson and Platnick, 1981; Page, 1990).

Their comment approaches the core of our present effort, but they describe a different purpose: "to create one or more area cladograms that all subtrees could agree with." The purpose of subtree analysis as implemented here is to specify the data relevant to cladistic biogeography—data that might conflict among themselves or not conflict as the case might be. Subtree analysis of their empirical example leads to a different result.

The cladogram for *Platynus* (fig. 33) includes 33 species, all but one of which (species 17) are endemic to one of eight areas in the region of Central America and the Antilles (species 17 is widespread in two of the areas). The cladogram reduces to Subtrees 1–4, which include eight informative nodes (fig. 34).

Subtrees 1–4 are similar to the subtrees of Page and Lydeard, but there are differences. Their four subtrees do not include Subtree 2. Their subtrees together include only five informative nodes (fig. 32), four of which are among the eight nodes of Subtrees 1–4 (the informative node of their subtree 1 is node 23 of Subtree 1; that of their subtree 2 is node

TABLE 21
Component and Three-item Matrices for Subtrees of Beetles[a]

	Nodes	Nodes of Subtrees							
	0 1 1 1 1 1 2 2 9 1 5 6 8 9 2 3	0 0 0 0 9 9 9 9	1 1 1 1 1 1	1 1 1 5 5 5	1 1 6 6	1 1 8 8	1 1 9 9	2 2	2 2 3 3
OG	0 0 0 0 0 0 0 0	0 0 0 0	0 0 0	0 0 0	0 0	0 0	0 0	0	0 0
LA	0 0 0 0 0 0 0 0	0 0 0 0	? ? 0	? ? 0	0 0	0 0	? 0	0	? 0
J	? ? ? 1 ? 0 ? ?	? ? ? ?	? ? ?	? ? ?	1 1	? ?	0 ?	?	? ?
CH	1 0 1 ? 1 ? ? 1	1 1 ? ?	0 ? ?	1 1 1	? ?	1 ?	? ?	?	1 1
SH	1 0 1 ? 1 ? ? 1	? ? 1 1	? 0 ?	1 1 1	? ?	? 1	? ?	?	1 1
EC	1 1 0 ? ? ? ? ?	? 1 ? 1	1 1 1	? 0 ?	? ?	? ?	? ?	?	? ?
WC	1 1 0 ? ? ? ? ?	1 ? 1 ?	1 1 1	0 ? ?	? ?	? ?	? ?	?	? ?
NoCA	? ? ? 1 1 1 1 0	? ? ? ?	? ? ?	? ? ?	1 ?	1 1	1 1	1	0 ?
CA	? ? ? 1 ? 1 1 ?	? ? ? ?	? ? ?	? ? ?	? 1	? ?	1 1	1	? ?

[a] Left, component matrix for eight informative nodes of four subtrees (fig. 34). Right, three-item matrix for eight informative nodes of four subtrees (fig. 34). Three-item matrix (right) is not derivable from component matrix (left). Symbols as in figure 32.

19 of Subtree 3; the two informative nodes of their subtree 3 are nodes 11 and 15 of Subtree 4). Their subtree 4 has one informative node, which is node 4 of the cladogram for *Platynus* (fig. 33); that node is paralogous and without equivalent in Subtrees 1–4 (fig. 34).

Parsimony analysis of a component matrix (table 21, left) for the eight nodes of Subtrees 1–4 (fig. 34) yields 96 trees, length 8, ci 100, ri 100. A strict consensus of the 96 trees has one node grouping all areas save the Lesser Antilles (LA). Search for a minimal tree reveals one tree (fig. 35), length 8, ci 100. Three-item analysis of the eight nodes yields 19 statements (table 21, right). Parsimony analysis of a uniformly weighted matrix yields 96 trees, length 19, ci 100, ri 100, with the same strict consensus and the same minimal tree. Parsimony analysis of a fractionally weighted (×4) matrix yields the same results, length 72, ci 100, ri 100. This minimal tree of four nodes (fig. 35) is offered as an exact result of parsimony analysis of the matrix for the paralogy-free fraction of data of all nodes of the cladogram for *Platynus*.

Of the 96 trees of shortest length for the component data, the number of nodes varies from three to six. There are 45 trees of six nodes, among which is the tree that Page and Lydeard found to have the fewest items of error (fig. 32, Area Cladogram). Items of error are determined by all nodes of the relevant tree, in this case all 28 nodes of the tree of figure 33. Of these 28 nodes, only 8 appear as informative nodes of Subtrees 1–4 (nodes 9, 11, 15, 16, 18, 19, 22, 23), leaving 20 nodes as paralogous. Items of error in this case are determined predominantly by paralogous nodes. The tree with fewest items of error is overresolved for the data (see also below).

MINIMALITY

The minimal tree for *Platynus* (fig. 35, Minimal Tree) has the interesting property of minimality relative to other trees of shortest length and fewer nodes (fig. 35, Non-Min Trees 1 and 2). Minimality, in the sense of "minimal tree" used here, does not mean merely fewer nodes, nor even less information (as measured by the number of three-item statements). Fewer nodes and fewer three-item statements are, nevertheless, indications of minimality in many cases.

Minimality relates to two interpretations of multiple branching (Nelson and Platnick, 1980) and also to basic procedures of three-item analysis (Nelson and Ladiges, 1991b: table 6, Example 6). The example offered by the latter authors is the simplest possible: A(BC) + A(DE) = A(BC)(DE). The example stated in words is that if B and C are related more closely than to A, and D and E are related more closely than to A, then the minimal tree describing both relationships is tree A(BC)(DE). There are many other nonminimal trees that describe the two relationships, for example tree A(BCDE). Why is the former (with two nodes) minimal relative to the latter (with only one node)?

A relevant observation is that an additional relationship, A(BD), changes the minimal tree from A(BC)(DE) to A(BCDE): A(BC) + A(DE) + A(BD) = A(BCDE). The latter tree, in other words, is determined by more rather than fewer data. Another observation is that no additional relationship—one different from but consistent with the three above—can change the minimal tree from A(BCDE) to A(BC)(DE). Addition of a relationship such as A(BE) leaves the minimal tree unchanged; addition of (BC)D changes the minimal tree to A((BC)DE). A final observation is that the polytomy of tree A(BC)(DE) is basal, in contrast to that of tree A(BCDE). According to

Nelson and Platnick (1980), a polytomy renders ambiguous any nodes distal to it. A tree with a basal polytomy is more ambiguous than a tree with a dichotomous basal node. A minimal tree, being least resolved, is most ambiguous relative to other trees of shortest length for a particular data matrix.

It was the view of Nelson and Platnick (1980) that minimal trees are implicit in systematics because of the necessity of working with trees that, in reflecting the current state of knowledge, are not perfectly resolved—trees that always contain multiple branchings. We have attempted to make explicit use of minimal trees in the cladistic analysis of geographic data—an enterprise that, relative to systematics in general, is more problematic because geographic data are generally less sufficient, in the current state of knowledge, for full resolution of a tree of areas.

Except for the minimal trees found here for North American fishes and tropical American beetles, how are systematic data for these two groups—which are well worked out by current standards—to be reliably understood in a geographic sense? The interpretations offered by Mayden (1988) and Page and Lydeard (1994) are not only overresolved for the data but, in the case of the fishes at least, in apparent conflict with them.

Consistency

The three studies, of which the results are analyzed above—midges, fishes, beetles—are typical of many efforts in modern systematics, which attempt to determine detailed relationships among organisms of a particular taxon. In a cladistic sense, success of effort is measured in number and reliabilty of nodes of cladograms offered as a result. Nodes are graphic representations of relationships and of taxa, too, whether taxa be formally named or not. The cladograms for midges, fishes, and beetles (figs. 8, 18, 33) are virtually fully resolved, and in that sense they are results of successful effort. Reliability of the nodes lies in another dimension of success, but that dimension embraces future discovery, with its possibiltites of confirmation and refutation.

A remarkable feature of subtree analysis of these cladograms is the near 100% consistency of their geographic data as shown by par-
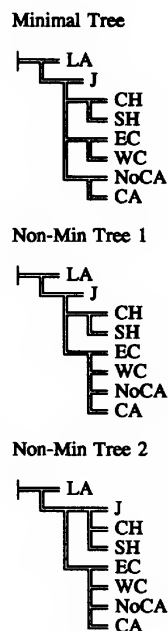
Fig. 35. Minimal and nonminimal trees, all of shortest length, resulting from parsimony analysis of matrices (table 21) for nodes of beetle subtrees (fig. 34). Symbols as in figure 32.

simony analysis of a matrix for the subtree nodes. Such consistency is not normally shown by results of parsimony analysis of systematic data. Nor is such consistency a reasonable expectation of any viewpoint based on an analogy between geographic and systematic data (Sober, 1988).

When such consistency is claimed in a particular case, the claim sometimes meets skepticism that such a result is obtainable without bias. An example is that of Edmunds (1981: fig. 6.16; comment in Nelson, 1982, 1984). In the 1979 symposium on vicariance biogeography, Edmunds presented five examples, one of scorpionflies (Mecoptera) and four of mayflies (Ephemeroptera). For each example, there are three genera, endemic to New Zealand, southeastern Australia, and southern South America, in which "the New Zealand genus is a sister group of the Australian-Magellanic [South American] pair." The pattern of distribution is one of the two types relevant to New Zealand noted by Brundin (see above): "a group in New Zealand is the sister group of a group occurring . . . in South America and Australia." Edmunds (1981:
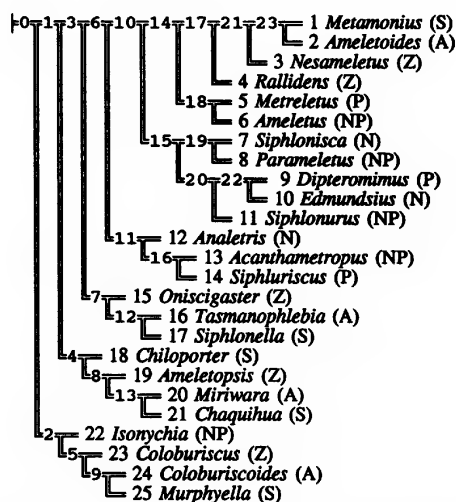
Fig. 36. Relationships of 25 genera of mayflies (after Edmunds, 1981: fig. 6.17), then of "family Siphlonuridae and part of family Oligonuridae" and subsequently subdivided into Siphlonuridae (node 6), Oniscigastridae (node 7), Ameletopsidae (node 4), and Coloburiscidae (node 2; e.g., Peters and Campbell, 1991). Symbols: A, Australia; N, Nearctic; P, Palearctic; S, South America; Z, New Zealand.
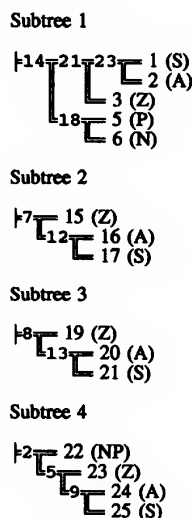

Fig. 37. Subtrees derived from cladogram of mayflies (fig. 36). Symbols as in figure 36.

293) commented, without other justification, that

> The question concerning the degree to which concordant cladograms were selected or "plucked" came up both in the auditorium and in private discussion. To allow persons to judge in the cases cited above I present a cladogram (figure 6.17) of the members of a highly paraphyletic group [of mayflies] considered to be one family at the time of analysis. Three other families—Baetidae, Caenidae, and Leptophlebidae—remain to be analyzed. The Nannochoristidae [Mecoptera] (figure 6.16) are a selected example suggesting that entire cool adapted lotic water communities were vicariated.

Edmunds' cladogram of mayflies (fig. 36) reduces to four subtrees (fig. 37), which for the southern areas duplicate his four examples. Remarkably, parsimony analysis of a matrix for all 24 nodes of the cladogram of mayflies (component data, table 22, above left) and of a matrix for the 7 informative nodes of the subtrees (component or three-item data, table 22, above right and below) yields one and the same tree with 100% consistency (the same as Subtree 1 of fig. 37). Edmunds' four examples and the consistency of their geograph-

ic data are, therefore, an unbiased representation of the cladogram of mayflies (fig. 36).

In this case, again by chance, the component data for all 24 nodes of the cladogram of mayflies (fig. 36) eliminate redundancy and paralogy such that, with parsimony analysis, a satisfactory result is achieved. Three-item analysis of the 24 nodes of the mayfly cladogram yields 1966 statements. Parsimony analysis of three-item matrices (uniformly and fractionally weighted $\times 8$) yields a different tree, S(A(Z(NP))), lengths 2964 and 22597, respectively, ci 66, ri 49. The different result with reduced consistency is attributable entirely to the effects of paralogy, as captured by the three-item data.

REVIEW OF MORRONE AND CARPENTER (1994)

For cladistic biogeography, Morrone and Carpenter (1994) reviewed available methods and computer implementations. They analyzed 10 data sets previously published, found that different methods give widely different results, and concluded that "current computer implementations of the methods remain unsatisfactory" (p. 114).

It is not necessary to apply subtree analysis to all 10 data sets in order to discover that most of the variability of their results stems not from different methods as such but rather

TABLE 22
Component and Three-Item Matrices for Mayflies[a]

	Nodes of Cladograms	Nodes
	0 0 0 0 0 0 0 0 0 0 1 1 1 1 1 1 1 1 1 1 2 2 2 2	0 0 1 1 1 2 2
	0 1 2 3 4 5 6 7 8 9 0 1 2 3 4 5 6 7 8 9 0 1 2 3	5 9 2 3 8 1 3
OG	0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0 0	0 0 0 0 0 0 0
A	1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 1 0 0 0 1 0 1	1 1 1 1 0 1 1
N	1 1 1 1 0 0 1 0 0 0 1 1 0 0 1 1 1 0 1 1 1 0 1 0	0 0 ? ? 1 0 0
P	1 1 1 1 0 0 1 0 0 0 1 1 0 0 1 1 1 0 1 1 1 0 1 0	0 0 ? ? 1 0 0
S	1 1 1 1 1 1 1 1 1 1 1 0 1 1 1 0 0 1 0 0 0 1 0 1	1 1 1 1 0 1 1
Z	1 1 1 1 1 1 1 1 1 1 0 1 0 0 0 1 0 0 1 0 0 1 0 0	1 0 0 0 0 1 0

	Nodes of Subtrees			
	1	2 3		4
	1 1 1 2 2 2 2	2 2 2	1 1	0 0 0 0 0 0 0
	8 8 8 1 1 1 1	3 3 3	2 3	5 5 5 5 9 9 9
OG	0 0 0 0 0 0 0	0 0 0	0 0	0 0 0 0 0 0 0
A	0 ? ? ? ? 1 1	1 1 1	1 1	1 1 ? ? 1 1 1
N	1 1 1 0 ? 0 ?	0 ? ?	? ?	0 ? 0 ? 0 ? ?
P	1 1 1 ? 0 ? 0	? 0 ?	? ?	? 0 ? 0 ? 0 ?
S	? 0 ? 1 1 ? ?	1 1 1	1 1	? ? 1 1 1 1 1
Z	? ? 0 1 1 1 1	? ? 0	0 0	1 1 1 1 ? ? 0

[a] Above left, component matrix for 24 nodes of clado-gram for mayflies (fig. 36). Above right, component ma-trix for seven informative nodes of four subtrees (fig. 37). Below, three-item matrix for seven informative nodes of four subtrees. Symbols: OG, Outgroup; A, Australia; N, Nearctic; P, Palearctic; S, South America; Z, New Zealand.
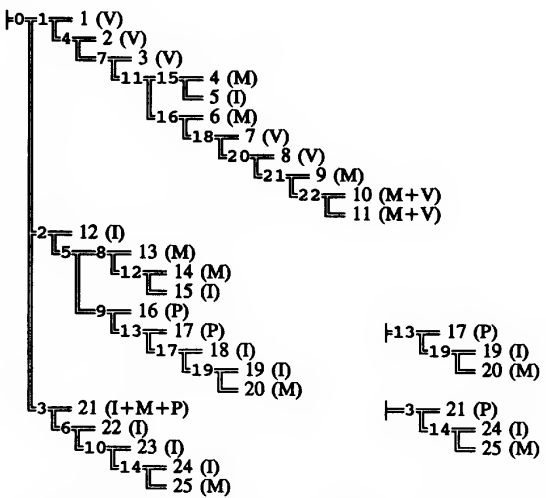

Fig. 38. **Left**, combination of cladograms of three genera of weevils (Insecta, Coleoptera, Curculion-idae) of southern South America. Cladograms (nodes): 1, *Germainiellus* (after Morrone, 1993a: fig. 52); 2, *Antarctobius* (after Morrone, 1992b: fig. 36); 3, *Falklandius* (after Morrone, 1992a: fig. 47). The corresponding cladograms of Morrone and Carpenter (1994: fig. 1) show fewer terminal taxa but preserve geographic data relevant to subtree analysis. **Right**, two subtrees derived from com-bined cladograms (left). Symbols: I, Islas Malvi-nas; M, Magellanic forest; P, Magellanic moor-land; V, Valdivian forest.

from the effects not only of paralogy as vari-ably captured by different methods, but also of missing data that cause programs to save overresolved trees. We apply subtree analysis to five of their data sets.

GERM (FIG. 38)

Three genera of weevils include species en-demic to the subantarctic domain of southern South America (Morrone, 1992a, 1992b, 1993a; summary in Morrone et al., 1994). Twenty-five species occur among four areas. The combined cladograms (fig. 38, left) re-duce to two identical subtrees of three areas only (fig. 38, right). The variable results found by Morrone and Carpenter (1994: tables 1, 4) include all 15 possible fully resolved trees for four areas. In a subsequent publication, Morrone and Anderson (1995) describe ad-ditional species and revise relationships within one genus (*Falklandius*, fig. 38, node 3), removing from it one species (fig. 38, spe-cies 25) and placing that species in a position

basal relative to all other species of the genus. These changes eliminate the second of the two subtrees and thereby diminish the results obtained by subtree analysis of cladograms of these organisms.

INDO (FIGS. 39, 40)

Ten groups of plant bugs include species endemic to Africa and Australasia (Schuh and Stonedahl, 1986). Forty-six species and spe-cies groups occur among 15 areas. The com-bined cladograms (fig. 39, left) reduce to 11 subtrees (fig. 39, right). A minimal tree (fig. 40, middle) derived from parsimony analysis of a component matrix for subtree nodes is similar to the hand resolution of Schuh and Stonedahl (fig. 40, above), which proves ov-erresolved for areas 3 and 7 (S India and N Burma). A minimal tree (fig. 40, below) de-rived from parsimony analysis of three-item matrices (no difference between uniform and fractional weighting) includes a division be-
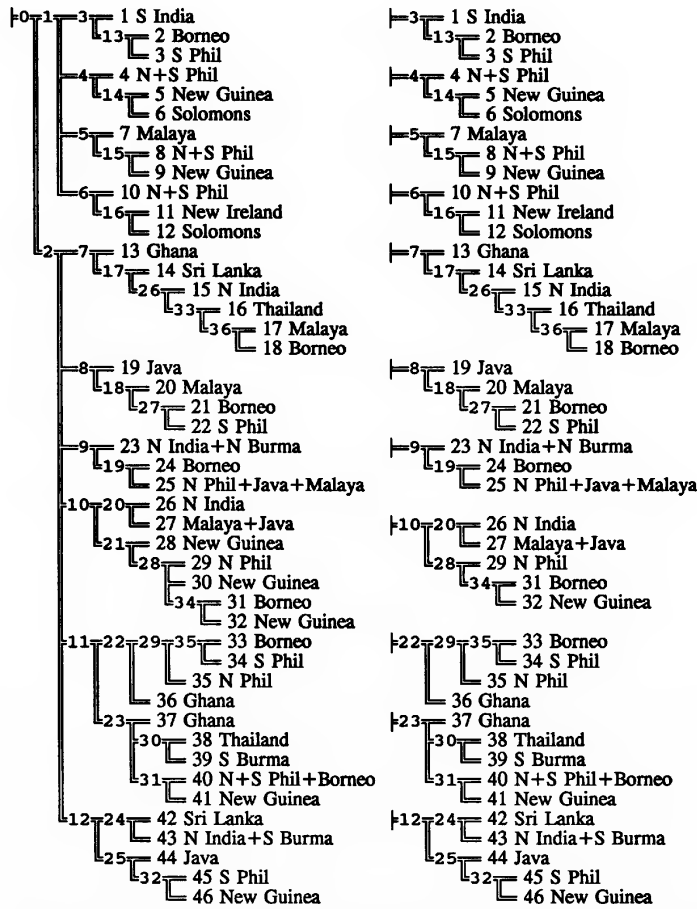
Left cladogram:

```
0─1─3─┬─ 1 S India
      13─┬─ 2 Borneo
         └─ 3 S Phil
   ─4─┬─ 4 N+S Phil
      14─┬─ 5 New Guinea
         └─ 6 Solomons
   ─5─┬─ 7 Malaya
      15─┬─ 8 N+S Phil
         └─ 9 New Guinea
   ─6─┬─ 10 N+S Phil
      16─┬─ 11 New Ireland
         └─ 12 Solomons
 2─7─┬─ 13 Ghana
     17─┬─ 14 Sri Lanka
        26─┬─ 15 N India
           33─┬─ 16 Thailand
              36─┬─ 17 Malaya
                 └─ 18 Borneo
   ─8─┬─ 19 Java
      18─┬─ 20 Malaya
         27─┬─ 21 Borneo
            └─ 22 S Phil
   ─9─┬─ 23 N India+N Burma
      19─┬─ 24 Borneo
         └─ 25 N Phil+Java+Malaya
 10─20─┬─ 26 N India
       └─ 27 Malaya+Java
    21─┬─ 28 New Guinea
       28─┬─ 29 N Phil
          ├─ 30 New Guinea
          34─┬─ 31 Borneo
             └─ 32 New Guinea
 11─22─29─35─┬─ 33 Borneo
             ├─ 34 S Phil
             └─ 35 N India
          └─ 36 Ghana
    23─┬─ 37 Ghana
       30─┬─ 38 Thailand
          └─ 39 S Burma
       31─┬─ 40 N+S Phil+Borneo
          └─ 41 New Guinea
 12─24─┬─ 42 Sri Lanka
       └─ 43 N India+S Burma
    25─┬─ 44 Java
       32─┬─ 45 S Phil
          └─ 46 New Guinea
```

Right cladograms (subtrees):

```
─3─┬─ 1 S India
   13─┬─ 2 Borneo
      └─ 3 S Phil
─4─┬─ 4 N+S Phil
   14─┬─ 5 New Guinea
      └─ 6 Solomons
─5─┬─ 7 Malaya
   15─┬─ 8 N+S Phil
      └─ 9 New Guinea
─6─┬─ 10 N+S Phil
   16─┬─ 11 New Ireland
      └─ 12 Solomons
─7─┬─ 13 Ghana
   17─┬─ 14 Sri Lanka
      26─┬─ 15 N India
         33─┬─ 16 Thailand
            36─┬─ 17 Malaya
               └─ 18 Borneo
─8─┬─ 19 Java
   18─┬─ 20 Malaya
      27─┬─ 21 Borneo
         └─ 22 S Phil
─9─┬─ 23 N India+N Burma
   19─┬─ 24 Borneo
      └─ 25 N Phil+Java+Malaya
10─20─┬─ 26 N India
      └─ 27 Malaya+Java
   28─┬─ 29 N Phil
      34─┬─ 31 Borneo
         └─ 32 New Guinea
22─29─35─┬─ 33 Borneo
         ├─ 34 S Phil
         └─ 35 N Phil
      └─ 36 Ghana
23─┬─ 37 Ghana
   30─┬─ 38 Thailand
      └─ 39 S Burma
   31─┬─ 40 N+S Phil+Borneo
      └─ 41 New Guinea
12─24─┬─ 42 Sri Lanka
      └─ 43 N India+S Burma
   25─┬─ 44 Java
      32─┬─ 45 S Phil
         └─ 46 New Guinea
```

Fig. 39. **Left**, combination of cladograms for 10 genera and species groups of plant bugs (Insecta, Heteroptera, Miridae) of the Indo-West Pacific Region (after Schuh and Stonedahl, 1986: figs. 5–7). Cladograms (nodes): 3, *Auricillocoris*; 4, *Ctypomiris* group; 5, *Sejanus ecnomious* species group; 6, *Leucophoroptera philippinensis* species group; 7, *Dioclerus*; 8, *Myiocarpus*; 9, *Mertila*; 10, *Harpedona*; 11, *Prodromus*; 12, *Thaumastomiris*. Other relevant nodes: 0, Miridae; 1, Phylinae; 2, Eccritotarsini. All widespread terminal taxa are species (4, 8, 10, 23, 25, 27, 40, 43). **Right**, 11 subtrees derived from combined cladograms (left). Abbreviations: Borneo, northern Borneo; N & S Phil, northern and southern Philippines; Thailand, northwestern Thailand.

tween areas of the Asian mainland (areas 2, 4–6, 9) and those of the Indo-Australian Archipelago (areas 8, 10–15). These three results (fig. 40, above, middle, below) are significantly consistent among themselves. Morrone and Carpenter (1994: table 6) find that most methods yield numerous trees (up to 5455) for these data.

MASA (FIG. 41)

Eighteen genera and one tribe of two subfamilies of wasps include species endemic to

seven continental areas of the world (Carpenter, 1993). The combined cladograms (fig. 41, left) reduce to 11 subtrees (fig. 41, right). Carpenter (1993) and Morrone and Carpenter (1994) noted that the geographic patterns shown by the two subfamilies (fig. 41, nodes 1 and 2, Masarinae and Polistinae, respectively) are different. The corresponding subtrees comprise two groups, one of five subtrees (fig. 41, subtrees including terminal taxa 1–14) and one of six subtrees (fig. 41, subtrees including terminal taxa 23–35). Parsimony analysis of a matrix for nodes of subtrees of

each group yields one minimal tree with 100% consistency: A(S(FNP)) for Masarinae, (NS)(ACFO) for Polistinae. Carpenter (1993: 153) construed the different patterns to reflect the relative age of origin of the two groups (Masarinae older, Polistinae younger): "North America and Australia may have been reached after breakup [of Gondwana] via dispersal in the paper wasps [Polistinae]." Morrone and Carpenter (1994: table 6) report varying results (up to 1000 trees) with various methods.

BIRD (FIGS. 42, 43)

Eight genera and species groups of birds of five families include species endemic to tropical South America (Cracraft, 1988). Forty species occur among nine areas. According to Cracraft, there are two geographic patterns. Four groups (fig. 42, left; nodes 5, 7, 17, 18) show variants of the pattern NE(SE(SW NW))—termed "Guinanan-Amazonian" (p. 222, pattern 1, his fig. 2)—with other areas variously represented (fig. 42: SEB, CA, CHO, NC, IM in node 5; SEB, CA, CHO, IM in node 7; IM in node 18). Area SEB occurs in two groups, once basally (node 5), once terminally (node 7). Four other groups (nodes 3, 6, 11, 12) show the pattern (NE NW)(SE SW)—termed "Trans-Amazonian" (p. 224, pattern 2, his fig. 3)—with no other areas included.

Cracraft treated the eight groups as subtrees, and represented their component data and species distributions in complete matrices: one matrix for each of the two assemblages of four groups and one matrix for all eight groups. Parsimony analysis of the complete matrix for the pattern 1 assemblage (nodes 5, 7, 17, 18, informative characters only) yields three trees (fig. 43, P1 Trees 1–3), length 20, ci 80, ri 85, all showing pattern 1: NE(SE(SW NW)). Parsimony analysis of the complete matrix for the pattern 2 assemblage (nodes 3, 6, 11, 12, informative characters only) yields one tree, length 8, ci 100, ri 100, showing pattern 2: (NE NW)(SE SW). Parsimony analysis of the complete matrix for both assemblages (informative characters only) yields five trees, length 35, ci 68, ri 68, all showing pattern 2: (NE NW)(SE SW). Cracraft (1988: 229) concluded that
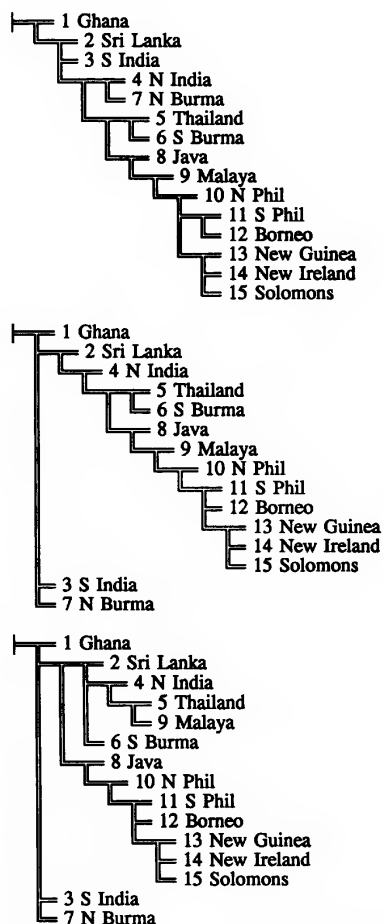
Fig. 40. Area cladograms for plant bugs. **Above,** after Schuh and Stonedahl (1986: fig. 9). **Middle,** minimal tree from parsimony analysis of component matrix for 21 nodes of 11 subtrees (fig. 39, right), length 23, ci 91, ri 93. **Below,** minimal tree from parsimony analysis of three-item matrix (87 statements) for 11 subtrees (fig. 39, right), length 99, ci 87, ri 86 (no difference between uniformly and fractionally weighted matrices).

Parsimony analysis has resulted, essentially, in a single historical hypothesis for the core areas of endemism within the Amazon basin [NE, NW, SE, SW]. These results are analogous in many respects to cases in which character distributions are ambiguous and narrowly favor one systematic hypothesis over another.

Cracraft (1988: 221, 230) found the result (of parsimony analysis of the complete matrix for both assemblages) unsatisfactory because

Current methods . . . are reductionist in the sense that they attempt to resolve multiple conflicting patterns
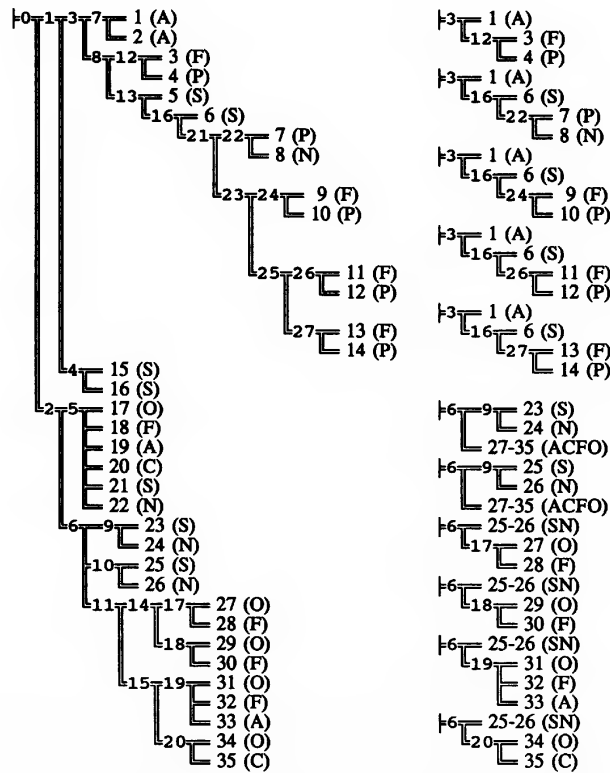
Fig. 41. **Left**, combination of cladograms for 18 genera and one tribe of two subfamilies of wasps (Insecta, Hymenoptera, Vespidae, Masarinae and Polistinae) with worldwide distribution (after Carpenter, 1993: figs. 7.6, 7.8). Cladograms (nodes): 1, Masarinae; 2, Polistinae. Widespread terminal taxa treated as nodes: *Ceramius* (12), *Jugurtia* (24), *Celonites* (26), *Quartinia* (27), *Polistes* (5), Epiponini (9), *Mischocyttarus* (10), *Polybioides* (14), *Belonogaster* (18), *Ropalidia* (19), *Parapolybia* (20). **Right**, 11 subtrees derived from combined cladograms (left). Symbols: A, Australia; C, Far East (China, Korea); F, Africa; N, North America; O, Oriental; P, Palearctic; S, South America.

across species-cladograms to a singular, less complex pattern. Parsimony analysis . . . is designed to reduce the complexity of multiple patterns to one (or more) most parsimonious hypothesis. Because it relies upon a questionable analogy to methods in systematics, however, biogeographic parsimony analysis has the potential to obscure the history of a biota rather than reveal it.

From this standpoint, therefore, a worthy ("nonreductionist") method should have found not one geographic pattern but two or more patterns (see below).

The combined cladograms (fig. 42, left) reduce to eight subtrees (fig. 42, right) corresponding to Cracraft's eight species groups and their geographic data. Many other groups of birds occur in these areas of South America, and those of the combined cladograms are only a sample. How the sampling was

accomplished is unexplained. Unlike Edmund's example of mayflies (see above), it is not possible to ascertain that the groups in the combined cladograms, and in the subtrees derived from them, are an unbiased representation of the entire cladogram of Amazonian birds. Rather, the groups seem to have been selected, or "plucked," for the purpose of illustrating Cracraft's argument and are hardly different from a hypothetical example concocted for the same purpose.

The subtrees may, nevertheless, be sorted into two classes, corresponding to Patterns 1 and 2. Parsimony analysis of a component matrix for nodes of Pattern 1 subtrees yields three trees (fig. 43, P1Trees 1–3), with a strict consensus showing a basal polytomy among areas 1, 2–3, 4, and 5–8 (fig. 43, Con3). Parsimony analysis of three-item matrices for
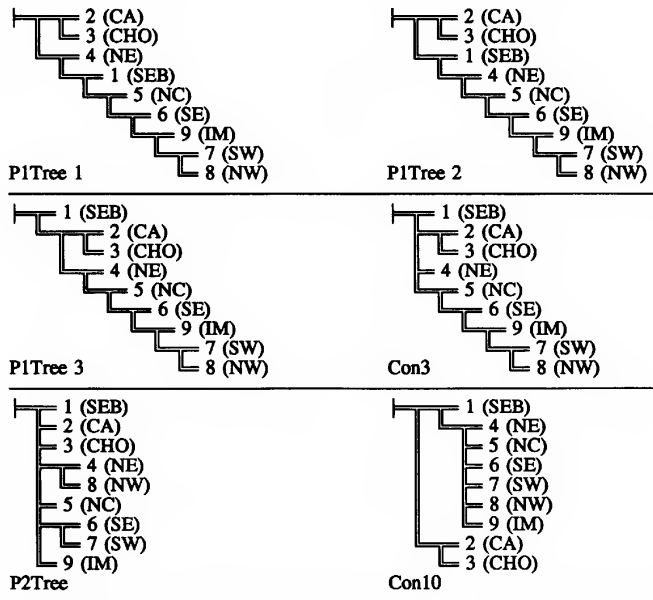
Fig. 42. **Left,** combination of cladograms for eight genera and species groups of birds of tropical South America (after Cracraft, 1988: figs. 2, 3). Cladograms (nodes): 3, *Psophia*; 5, *Pionopsitta*; 6, *Pionites*; 7, *Selenidera*; 11, *Lanio*; 12, *Pipra*; 17, *Pteroglossus viridis* group; 18, *Pteroglossus bitorquatus* group. Other relevant nodes: 1, Psittacidae; 2, Ramphastidae; 4, Passeriformes; 8, *Pteroglossus*. **Right,** eight subtrees derived from combined cladograms (left). Symbols: CA, Central America; CHO, Choco; IM, Imeri; NC, Nechi; NE & NW, northeasten and northwestern Amazonia; SE & SW, southeastern and southwestern Amazonia; SEB, southeastern Brazil.

nodes of Pattern 1 subtrees yields either (uniformly weighted matrix) one tree (fig. 43, P1 Tree 3) or (fractionally weighted matrix × 12) two trees (fig. 43, P1 Trees 2 and 3). Parsimony analysis of a component matrix for nodes of Pattern 2 subtrees yields one minimal tree (fig. 43, P2 Tree), and three-item matrices yield the same. Parsimony analysis of a component matrix for nodes of all eight subtrees yields 10 trees, all showing Pattern 2, with a strict consensus of only three nodes (fig. 43, Con10). Parsimony analysis of a three-item matrix for nodes of all eight subtrees yields either (uniformly weighted matrix) one tree (fig. 43, P1 Tree 3) or (fractionally weighted × 12 matrix) two trees (fig. 43, P1 Trees 2 and 3).

The results of parsimony analysis of a component matrix for nodes of all subtrees are hardly different from those obtained by Cracraft (the corresponding matrices are nearly the same). The results of parsimony analysis of a three-item matrix for nodes of all subtrees differ, showing Pattern 1 rather than Pattern 2. That difference exemplifies Cracraft's concern that ambiguous characters, depending on the method used, might "narrowly favor one . . . hypothesis over another," but the results jointly recover both Patterns 1 and 2 from the entire data and in that sense function in this case as the "nonreductionist methodology" that Cracraft sought for but did not attain:

Biogeographic methodology needs to develop analytical techniques in order that complex historical pat-

Fig. 43. Results of parsimony analysis of matrices for subtrees derived from combined cladograms of eight groups of birds (fig. 42). P1Trees 1–3, of component matrix for subtrees (13 nodes only) corresponding to Pattern 1 (subtrees with basal nodes 5, 7, 17, 18), length 16, ci 81, ri 87. Con3, strict consensus of P1Trees 1–3. P2Tree, minimal tree, from parsimony analysis of component matrix for nodes of subtrees (eight nodes only) that correspond to Pattern 2 (subtrees with basal nodes 3, 6, 11, 21), length 8, ci 100, ri 100. Con10, strict consensus of 10 trees, from parsimony analysis of component matrix for 21 nodes of eight subtrees, length 31, ci 67, ri 68. Parsimony analysis of three-item matrices (129 statements) for Pattern-1 subtrees (eight nodes only) yields either P1Tree3 (uniformly weighted matrix), length 146, ci 88, ri 86, or P1Trees 2 and 3 (fractionally weighted ×12 matrix), length 1494, ci 89, ri 88. Parsimony analysis of three-item matrices (145 statements) for all 21 nodes of subtrees yields either P1Tree 3 (uniformly weighted matrix), length 174, ci 83, ri 80, or P1Trees 2 and 3 (fractionally weighted ×12 matrix), length 1830, ci 83, ri 80.

terns are not concealed by estimates that are very much less complex (p. 233).

Morrone and Carpenter (1994: table 5) found that some methods yield only one tree, and most methods two or three trees, for these data. Seven different trees were found (their fig. 7), all, curiously, showing Pattern 1: NE(SE(SW NW)).

LIST (FIG. 44)

Two genera and one species group of a third genus of weevils and one species group of asters include species endemic to southern South America (Morrone, 1993b; Lanteri, unpub.; Anderberg and Freire, 1991). Forty-two terminal taxa occur among four areas. The combined cladograms reduce to five subtrees, which conflict among themselves. Parsimony analysis of a component matrix for

nodes of all subtrees yields two trees, B(D(AC)) and B(A(CD)), length 8, ci 75, ri 66, with strict consensus B(ACD). Parsimony analysis of three-item matrices (eight statements) for all nodes of subtrees yields the same, length 10, ci 80, ri 75 (no difference between uniformly and fractionally weighted matrices). Morrone (1993b: fig. 6) found one tree, A(B(CD)). With a variety of methods, Morrone and Carpenter (1994: table 6) found either one tree or three trees. In their analyses of each of the four groups, they found several trees, usually all of the 15 possible resolved trees for four areas. Subtree analysis, in contrast, yields a single tree for each of the four groups: subtrees 1 and 2 combine with 100% consistency as D(A(BC)), which conflicts with the other three subtrees, which are 100% consistent and combine as B(A(CD)). If conflict is reckoned as evidence of more than one
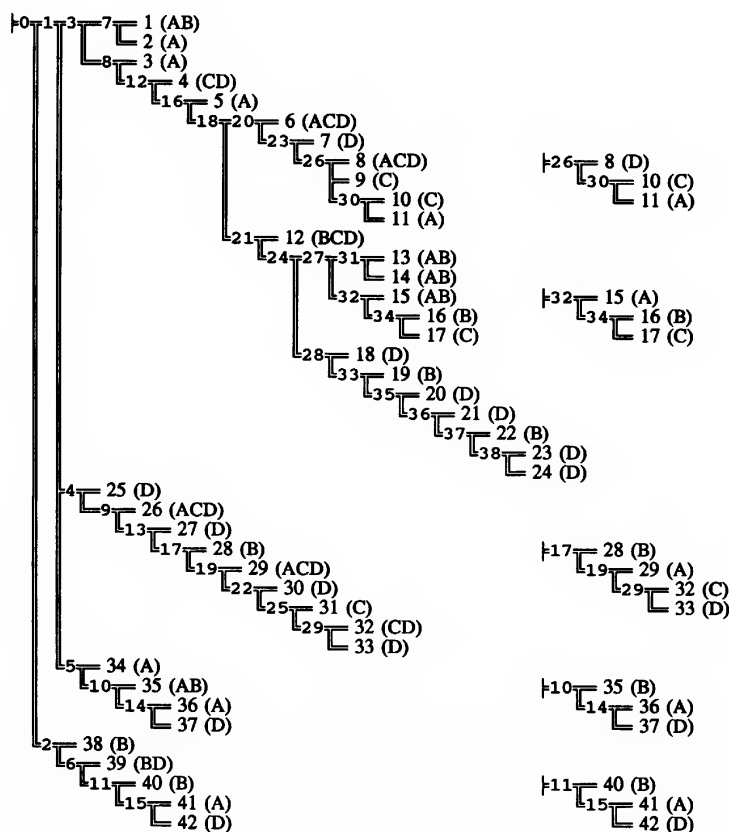
Left diagram:

```
0-1-3-7- 1 (AB)
         2 (A)
       8- 3 (A)
         12- 4 (CD)
            16- 5 (A)
               18-20- 6 (ACD)
                     23- 7 (D)
                        26- 8 (ACD)
                           9 (C)
                          30- 10 (C)
                             11 (A)
               21- 12 (BCD)
                  24-27-31- 13 (AB)
                           14 (AB)
                          32- 15 (AB)
                             34- 16 (B)
                                17 (C)
                  28- 18 (D)
                     33- 19 (B)
                        35- 20 (D)
                           36- 21 (D)
                              37- 22 (B)
                                 38- 23 (D)
                                    24 (D)
   4- 25 (D)
     9- 26 (ACD)
       13- 27 (D)
          17- 28 (B)
             19- 29 (ACD)
                22- 30 (D)
                   25- 31 (C)
                      29- 32 (CD)
                         33 (D)
   5- 34 (A)
     10- 35 (AB)
        14- 36 (A)
           37 (D)
   2- 38 (B)
     6- 39 (BD)
       11- 40 (B)
          15- 41 (A)
             42 (D)
```

Right subtrees:

```
26- 8 (D)
  30- 10 (C)
     11 (A)

32- 15 (A)
  34- 16 (B)
     17 (C)

17- 28 (B)
  19- 29 (A)
     29- 32 (C)
        33 (D)

10- 35 (B)
  14- 36 (A)
     37 (D)

11- 40 (B)
  15- 41 (A)
     42 (D)
```

Fig. 44. **Left**, combination of two genera and one species group of weevils (Insecta, Coleoptera, Curculionidae) and one genus group of asters (Asteraceae, Gnaphaliae) of southern South America. Cladogram (nodes): 2, *Lucilia* genus group (after Anderberg and Freire, 1991); 3, *Listroderes* (after Morrone 1993b); 4, *Naupactus taeniatulus* species group (after Lanteri, personal commun. to Morrone, 1993b: 402); 5, *Hyperoides* (after Morrone, 1993c). Other relevant node: 1, Curculionidae. The corresponding cladograms of Morrone and Carpenter (1994: fig. 3) show fewer terminal taxa but preserve grographic data relevant to subtree analysis. **Right**, five subtrees derived from combined cladograms (left). Symbols: A, central Chile; B, subantarctic; C, central Argentina; D, Chaco.

---

geographic pattern, then in these data there is evidence of two patterns.

THE PAST AND THE FUTURE

For cladistic biogeography, the facts of geographic distribution of organisms are coherent patterns related to, and explained by, historical processes of geographic change. This view had been epitomized in the slogans of Croizat (1964: 857–858) that "dispersal forever repeats" and that "earth and life evolve together." During recent decades this view was strengthened by developments both in geology and in biology—by the revival of continental drift (plate tectonics) and by the

rejuvenation of systematics (cladistics). These developments did not in themselves render more coherent the facts of geographic distribution of organisms, but they heightened expectations that such discovery was within reach of empirical investigation.

For a long time, the expectations had been in existence. Early in the modern history of cladistics they were reformulated (Hennig, 1960, 1966) and given significant impetus (Brundin, 1966). Subsequent developments within cladistics offered the hope that geographic data, when associated with nodes of cladograms generally and when analyzed by the exact methods of parsimony analysis, would prove coherent—even convincing to

other, ecologically oriented, biogeographers and to biologists in general. This hope, still persistent today, seems to have been realized only to a limited degree. It is doubtful, for example, that the accumulated findings of cladistic biogeography of the last two decades have proven any more convincing than those of Croizat of the previous few decades (bibliography in Craw and Gibbs, 1984; comment by Seberg, 1986; Platnick and Nelson, 1989; Mayden, 1991; Nelson, 1994).

Subtrees simplify the cladistic analysis of geographic data. Subtree analysis does so by identifying paralogous nodes so that geographic data need not be associated with paralogous nodes. In this respect the rationale of subtree analysis is at variance with points of view, including our own at times, that have surfaced during the recent history of cladistic biogeography. In retrospect, these points of view seem tacitly to assume that nodes are composed (in part) of geographic data and that troubling variation of geographic data from node to node is random among cladograms in general. The contrasting assumption of subtree analysis is that troubling variation among geographic data that might be associated with some nodes is merely the effect of paralogy, which is nonrandom and increases basally in cladograms in general.

Subtree analysis captures some, at times many, of the routine practices of biogeographic analysis, for example the practice, seemingly universal, of comparing cladograms of different groups of organisms for the purpose of identifying such common patterns—or such different patterns—as might be present in the associated geographic data (Grande, 1994).

No one would argue that there is more than one tree of life. The rationale of subtree analysis addresses the question of how that one tree—to the degree that it is currently known—is best subdivided for the purposes of comparison and geographic analysis. To our knowledge, this question has never before been addressed, nor has any algorithm (rationale) been previously offered as its answer. Such an algorithm seems required if cladistic biogeography is to have a rational basis. The lack of such basis perhaps explains the frustration sometimes expressed, for good reason, with results of current methods of analysis (e.g., by Morrone and Carpenter, 1994). With such basis, one may evaluate, even improve upon, the division of that one tree into two or more parts (subtrees) that might be assumed comparable by whomsoever.

Heretofore, the division, accomplished arbitrarily in all cases known to us, has been into groups (taxa) independent in the sense that they have no nodes in common. Subtree analysis imposes no such requirement, and one or more nonterminal node may appear as an element common to two or more subtrees. This novel feature of subtree analysis may appear counterintuitive because independence of subtrees is subsequently assumed under the rationale of parsimony analysis.

Here we forego discussion of whether parsimony analysis (of areas) is essential to, or ultimately meaningful for, cladistic biogeography. Relevant, nevertheless, is that the subtree algorithm requires no computer program and can be implemented by hand even for complex cladograms. The resulting subtrees tend to be simple, and their meaning and consistency (or mutual conflict) are generally evident without parsimony analysis.

Subtree analysis offers apparent advantages over existing procedures, as is evident in the above analyses. Some data that seemed obscure (e.g., the GERM data reviewed by Morrone and Carpenter) are rendered clear and simple. Other data that seemed clear enough but otherwise suspect (e.g., those of Edmunds) are given rational justification. These appealing features seem general properties of subtree analysis, which might not be the ultimate solution for cladistic biogeography but is possibly a step in that direction.

SUMMARY

1. Nodes of cladograms of organisms are not directly informative of geographic relationships that might exist between organisms. For nodes to be informative, geographic data must first be deliberately associated with them. Nodes (taxa) relating organisms in geographic areas that overlap are deemed "paralogous" (by analogy with the "paralogy" of molecular biology). Geographic data need not be associated with paralogous nodes.

2. Subtree analysis is a novel method of

potential value in cladistic biogeography. It proceeds by reducing one more or less complex cladogram to a one or more subtree that is paralogy free in the geographic sense. Geographic data are then associated with informative nodes of each subtree. Data so associated appear to be the only data actually relevant to cladistic biogeography.

3. Subtree analysis accords with the general practice of comparing cladograms for two or more taxa for the purpose of cladistic analysis of geographic relationship. Because cladograms for two or more groups of organisms are parts of the one cladogram embracing all of life, subtree analysis provides a rationale for subdividing that one cladogram for that purpose.

4. An algorithm for subtree analysis was developed, implemented in a preliminary MS-DOS program, and applied to the benchmark studies in cladistic biogeography of Brundin (1966) and Mayden (1988), as well as to cladograms in studies reviewed by Craw (1989), Page and Lydeard (1994), and Morrone and Carpenter (1994).

5. Data associated with nodes of subtrees were represented in matrices of two kinds (component and three item) for the purpose of parsimony analysis, which generally yielded the same trees of high consistency, approaching or reaching 100%. Problems previously encountered with parsimony analysis

of geographic data, as reported, for example, in the review by Morrone and Carpenter (1994), are largely the effects of geographic paralogy and disappear with subtree analysis.

6. Representation in matrices of geographic data (associated with nodes of subtrees) usually entails numerous missing-data entries. Missing data commonly permit current programs (Hennig86, PAUP) to save over-resolved trees (with spurious nodes) that prevent straightforward realization of an informative area cladogram. In such cases, an informative result is generally obtainable by manual collapse of spurious nodes and more basal placement, by trial and error, of single areas and combinations of them.

7. Geographic data have sometimes been associated with widespread species and represented in a matrix for parsimony analysis. Such data apparently add little or nothing (except lower consistency and spurious resolution) to results achievable by parsimony analysis of a matrix for data associated with nodes only.

ACKNOWLEDGMENTS

REFERENCES

Anderberg, A. A., and E. Freire
    1991.   A cladistic and biogeographic analysis of the *Lucilia* group (Asteraceae, Gnaphaliae). Bot. J. Linn. Soc. 106: 173–198.
Breure, A. S. H.
    1979.   Systematics, phylogeny and zoogeography of Bulimulinae (Mollusca). Zool. Verh. (Leiden) 168: 215 pp.
Brooks, D. R.
    1990.   Parsimony analysis in historical biogeography and coevolution: methodological and theoretical update. Syst. Zool. 39: 14–30.
Brundin, L.
    1963.   Limnic Diptera in the bearings on the problem of transantarctic faunal connections. *In* J. L. Gressit (ed.), Pacific Basin biogeography, pp. 425–434. Honolulu: Bishop Museum Press.
    1965.   On the real nature of transantarctic relationships. Evolution 19: 496–505.
    1966.   Transantarctic relationships and their significance, as evidenced by chironomid midges with a monograph of the subfamilies Podonominae and Aphroteninae and the austral Heptagyiae. Kungliga Svenska Vetenskapsakademiens Handligar, Fjärde Serien, 11(1): 472 pp.
    1967.   Insects and the problem of austral disjunctive distributions. Annu. Rev. Entomol. 12: 149–168.
    1970.   Antarctic land faunas and their history.

In M. W. Holdgate (ed.), Antarctic ecology 1: 41–53. London: Scientific Committee for Antarctic Research, Academic Press.

1972. Phylogenetics and biogeography. Syst. Zool. 21: 69–79.

1974. Fifty years limnic zoogeography. Mitt. Int. Ver. Limnol. 20: 287–300.

1975. Circum-antarctic distribution patterns and continental drift. *In* Thème No. 1: Biogéographie et liaisons intercontinentales au cours du Mésozoïque. XVIIe Congrès Internationale de Zoologie, Monaco – 25–30 Septembre 1972. Mém. Mus. Natl. Hist. Nat., n. sér. A, Zool., 88: 19–27.

1976. A Neocomian chironomid and Podonominae-Aphroteniinae (Diptera) in the light of phylogenetics and biogeography. Zool. Scr. 5: 139–160.

1981. Croizat's panbiogeography versus phylogenetic biogeography. *In* G. Nelson and D. E. Rosen (eds.), Vicariance biogeography: a critique, pp. 94–138. New York: Columbia Univ. Press.

1988. Phylogenetic biogeography. *In* A. A. Myers and P. S. Giller (eds.), Analytical biogeography: an integrated approach to the study of animal and plant distributions, pp. 343–369. London: Chapman and Hall.

1993. From Grimsgöl to Gondwanaland – half a century with chironomids. Cladistics 9: 358–367.

Carpenter, J. M.
1993. Biogeographic patterns in the Vespidae (Hymenoptera): two views of Africa and South America. *In* P. Goldblatt (ed.), Biological relationships between Africa and South America, pp. 139–155. New Haven, CT: Yale Univ. Press.

Cracraft, J.
1988. Deep-history biogeography: retrieving the historical patterns of evolving continental biotas. Syst. Zool. 37: 221–236.

Craw, R. C.
1989. New Zealand biogeography: a panbiogeographic approach. New Zealand J. Zool. 16: 527–547.

Craw, R. C., and G. W. Gibbs (eds.)
1984. Croizat's panbiogeography & principia botanica: search for a novel biological synthesis. Tuatara 27: 1–75.

Croizat, L.
1964. Space, time, form: the biological synthesis. Caracas: published by the author.

Edmunds, G. F., Jr.
1981. Discussion. *In* G. Nelson and D. E. Ro-

sen (eds.), Vicariance biogeogrphy: a critique, pp. 287–297. New York: Columbia Univ. Press.

Farris, J. S.
1988. Hennig86, version 1.5. Published by the author.

Funk, V. A., and D. R. Brooks
1990. Phylogenetic systematics as the basis of comparative biology. Smithson. Contrib. Bot. 73: iii + 45 pp.

Grande, L.
1994. Repeating patterns in nature, predictability and "impact" in science. *In* L. Grande and O. Rieppel (eds.), Interpreting the hierarchy in nature: from systematic patterns to evolutionary process theories, pp. 61–84. San Diego: Academic Press.

Hart, J. A.
1987. A cladistic analysis of conifers: preliminary results. J. Arnold Arbor., Harv. Univ. 68: 269–307.

Hennig, W.
1960. Die Dipteren-Fauna von Neuseeland als systematisches und tiergeographisches Problem. Beitr. Entomol. 10: 221–329.

1966. The Diptera fauna of New Zealand as a problem in systematics and zoogeography. Pac. Insects Monogr. 9: 81 pp.

Humphries, C. J.
1992. Cladistic biogeography. *In* P.L. Forey et al., Cladistics: a practical course in systematics. Syst. Assoc. Pub. 10: 137–159. Oxford: Oxford Univ. Press.

Humphries, C. J., and L. R. Parenti
1986. Cladistic biogeography. Oxford: Oxford Univ. Press.

Humphries, C. J., M. Cox, and E. S. Nielsen
1986. *Nothofagus* and its parasites: a cladistic approach to coevolution. *In* A. R. Stone and D. L. Hawksworth (eds.), Coevolution and systematics, pp. 55–76. Oxford: Oxford Univ. Press.

Humphries, C. J., P. Y. Ladiges, M. Roos, and M. Zandee
1988. Cladistic biogeography. *In* A. A. Myers and P. S. Giller (eds.), Analytical biogeography: an integrated approach to the study of animal and plant distributions, pp. 371–404. London: Chapman and Hall.

Legendre, P.
1990. Quantitative methods and biogeographic analysis. *In* D. J. Garbary and R. R. South (eds.), Evolutionary biogeography of the marine algae of the North Atlantic. NATO ASI Series Vol. G22, pp. 9–34. Berlin: Springer-Verlag.

Liebherr, J. K.
1988.   General patterns in West Indian insects, and graphical biogeographic analysis of some circum-Caribbean *Platynus* beetles (Carabidae). Syst. Zool. 37: 385–409.

Mayden, R. L.
1988.   Vicariance biogeography, parsimony, and evolution in North American freshwater fishes. Syst. Zool. 37: 329–355.
1991.   The wilderness of panbiogeography: a synthesis of space, time, and form? Ibid. 40: 503–519.

Mivart, St. G.
1877.   On the axial skeleton of the Struthionidae. Trans. Zool. Soc. London 10: 1–52.

Morrone, J. J.
1992a.   Revisión sistemática, análisis cladístico y biogeografía histórica de los géneros *Falklandius* Enderlein y *Lanteriella* gen. nov. (Coleoptera: Curculionidae). Acta Entomol. Chilena 17: 157–174.
1992b.   Revisión sistemática y análisis cladístico del género *Antarctobius* Fairmaire (Coleoptera: Curculionidae). Neotrópica (La Plata) 38: 3–20.
1993a.   Revisión sistemática de un nuevo género de Rhytirrhinini (Coleoptera: Curculionidae), con un análisis biogeográfico del domino subantártico. Bol. Soc. Biol. Concepción 64: 121–145.
1993b.   Cladistic and biogeographic analyses of the weevil genus *Listroderes* Schoenherr (Coleoptera: Curculionidae). Cladistics 9: 397–411.
1993c.   Revisión sistemática del género *Hyperoides* Marshall (Coleoptera: Curculionidae). Neotrópica (La Plata) 39: 17–26.

Morrone, J. J., and R. S. Anderson
1995.   The *Falklandius* generic group: cladistic analysis with description of new taxa (Coleoptera: Curculionidae: Rhytirrhinini). Am. Mus. Novitates 3121: 14 pp.

Morrone, J. J., and J. M. Carpenter
1994.   In search of a method for cladistic biogeography: An empirical comparison of component analysis, Brooks parsimony analysis, and three-area statements. Cladistics 10: 99–153.

Morrone, J. J., S. R. Juñent, and J. V. Crisci
1994.   South American beetles. Natl. Geogr. Res. Explor. 10: 104–115.

Nelson, G.
1982.   Cladistique et biogéographie. C.R. Soc. Biogéogr. 58: 75–94.
1984.   Cladistics and biogeography. *In* T. Duncan and T. F. Stuessy (eds.), Cladistics: perspectives on the reconstruction of evolutionary history, pp. 273–293. New York: Columbia Univ. Press.
1992.   Reply to Harvey. Cladistics 8: 355–360.
1994.   Homology and systematics. *In* B. K. Hall (ed.), Homology: the hierarchical basis of comparative biology, pp 101–149. San Diego, CA: Academic Press.

Nelson, G., and P. Y. Ladiges
1991a.   Standard assumptions for biogeographic analysis. Australian Syst. Bot. 4: 41–58 (Addendum: 5: 247).
1991b.   Three-area statements: standard assumptions for biogeographic analysis. Syst. Zool. 40: 470–485.
1993.   Missing data and three-item analysis. Cladistics 9: 111–113.
1995.   TAX: MSDos computer programs for systematics. New York and Melbourne: Published by the authors.

Nelson, G., and N. I. Platnick
1980.   Multiple branching in cladograms: two interpretations. Syst. Zool. 29: 86–91.
1981.   Systematics and biogeography: Cladistics and vicariance. New York: Columbia Univ. Press.

Page, R. D. M.
1988.   Quantitative cladistic biogeography: constructing and comparing area cladograms. Syst. Zool. 37: 254–270.
1990.   Component analysis: a valiant failure? Cladistics 6: 119–136.
1993.   Genes, organisms, and areas: the problem of multiple lineages. Syst. Biol. 42: 77–84.
1994.   Maps between trees and cladistic analysis of historical associations among genes, organisms, and areas. Ibid. 43: 58–77.

Page, R. D. M., and C. Lydeard
1994.   Towards a cladistic biogeography of the Caribbean. Cladistics 10: 21–41.

Peters, W. L., and I. C. Campbell
1991.   Ephemeroptera. *In* CSIRO (ed.), The insects of Australia: a textbook for students and research workers (2nd ed.) 1: 279–293. Carlton: Melbourne Univ. Press.

Platnick, N. I., and G. Nelson
1989.   Spanning-tree biogeography: shortcut, detour, or dead-end? Syst. Zool. 37: 410–419.

Platnick, N. I., C. E. Griswold, and J. A. Coddington
1991.   On missing entries in cladistic analysis. Cladistics 7: 337–343.

Schuh, R. T., and G. M. Stonedahl

1986. Historical biogeography in the Indo-Pacific: a cladistic approach. Cladistics 2: 337–355.

Seberg, O.
1986. A critique of the theory and method of panbiogeography. Syst. Zool. 35: 369–380.
1988. Taxonomy, phylogeny, and biogeography of the genus *Oreobolus* R. Br. (Cyperaceae), with comments on the biogeography of the South Pacific continents. Bot. J. Linn. Soc. 96: 119–195.

Sibley, G., and J. Ahlquist
1987. Avian phylogeny reconstructed from comparisons of the genetic material DNA. *In* C. Patterson (ed.), Molecules and morphology in evolution: conflict or compromise?, pp. 95-121. Cambridge: Cambridge Univ. Press.

Sober, E.
1988. The conceptual relationship of cladistic phylogenetics and vicariance biogeography. Syst. Zool. 37: 245–253.

Swofford, D. L.
1993. PAUP: phylogenetic analysis using parsimony, version 3.1.1. Washington, D.C: Smithsonian Institution.

Wiley, E. O.
1987. Methods in vicariance biogeography. *In* P. Hovenkamp *et al.* (eds.), Systematics and evolution: a matter of diversity, pp. 283-306. Utrecht: Utrecht Univ.
1988a. Parsimony analysis and vicariance biogeography. Syst. Zool. 37: 271–290.
1988b. Vicariance biogeography. Annu. Rev. Ecol. Syst. 19:513–542.

Wiley, E. O., D. Siegel-Causey, D. R. Brooks, and V. A. Funk
1991. The compleat cladist: a primer of phylogenetic procedures. Univ. Kansas Mus. Nat. Hist. Spec. Publ. 19: x + 158 pp.

Zandee, M., and M. C. Roos
1987. Component-compatibility in historical biogeography. Cladistics 3: 305–332.